Grundlehren der mathematischen Wissenschaften 303

A Series of Comprehensive Studies in Mathematics

Ronald A. DeVore George G. Lorentz

Constructive Approximation

With 5 Figures

Springer-Verlag
Berlin Heidelberg New York
London Paris Tokyo
Hong Kong Barcelona
Budapest

Ronald A. DeVore
Department of Mathematics
University of South Carolina
Columbia, SC 29208, USA

George G. Lorentz
Department of Mathematics
The University of Texas at Austin
Austin, TX 78712, USA

Mathematics Subject Classification (1991): Primary: 41xx, 42xx;
Secondary: 46xx, 47xx, 26Cxx, 26Dxx

ISBN 3-540-50627-6 Springer-Verlag Berlin Heidelberg New York
ISBN 0-387-50627-6 Springer-Verlag New York Berlin Heidelberg

Library of Congress Cataloging-in-Publication Data
DeVore, Ronald A.
Constructive approximation: polynomials and splines approximation /
Ronald A. DeVore, George G. Lorentz.
p. cm. -- (Grundlehren der mathematischen Wissenschaften; 303)
Includes bibliographical references and index.
ISBN 3-540-50627-6 (Berlin: acid-free).
ISBN 0-387-50627-6 (New York: acid-free)
1. Approximation theory. I. Lorentz, G. G. II. Title. III. Series
QA221.D44 1993 511'.42–dc20 93-18420 CIP

Printed in the United States of America

Typesetting: Data conversion by Springer-Verlag in using a
Springer TEX macro package
41/3140-5 4 3 2 1 0 Printed on acid-free paper

Preface

The present book deals with some basic problems of Approximation Theory: with properties of polynomials and splines, with approximation by polynomials, splines, linear operators. It also provides the necessary material about different function spaces. In some sense, this is a modern version of the corresponding parts of the book of one of us (Lorentz [A-1966]).

We have tried to give a complete exposition of the main, basic theorems of the theory, without going into too much detail, treating the most general cases or discussing very special problems. There are essential limitations: this is a book about approximation of functions of one real variable. But Approximation Theory of functions of several real or of complex variables would require new books. Very little is given on interpolation. But even with these restrictions, proofs of some deep and important results, like Korneichuk's theorems about approximation in Lipschitz spaces, could not be included. Another book, with different authors, "Constructive Approximation, Advanced Problems", is in preparation.

There is an extensive bibliography, which can be used also as an Author Index: each paper is supplied with references for the page or pages where it has been used.

We are indebted to several mathematicians who have supported us with valuable advice: Berens, Braess, Chui, Ditzian, Jetter, Leviatan, Stens, Totik, Xu and others. But our particular thanks are to C. de Boor and M. v. Golitschek, for their important ideas and proofs. We are deeply grateful to the Springer-Verlag for including this book in their excellent Grundlehren series.

The authors

Contents

Chapter 1. Theorems of Weierstrass

§ 1. Basic Notions

The main problem of approximation consists in finding for a complicated function f from a large space X a close-by, simple function ϕ from a small subset Φ of X. There are three elements here. The large space X is usually a normed space, such as C, L_p or one of the other Banach spaces of functions. The distance from ϕ to f can then be measured by the norm $\|f - \phi\|$ in X. Finally, we have to define the special functions of Φ. There are many possibilities, but the following three are basic in the theory.

For functions on a compact interval $[a, b]$ one often chooses $\Phi = \mathcal{P}_n$ to be the space of all *algebraic polynomials* P of degree $\leq n$,

$$P(x) := P_n(x) = \sum_{k=0}^{n} a_k x^k . \tag{1.1}$$

For the circle $\mathbb{T}$ (which we will usually understand to be $\mathbb{R}$ with the identification of points modulo 2π), natural is the class $\mathcal{T}_n$ of *trigonometric polynomials* of degree $\leq n$,

$$T(x) := T_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} (a_k \cos kx + b_k \sin kx) . \tag{1.2}$$

Finally, the third important class is the *piecewise polynomials;* they are called *splines* (we shall meet them in Chapter 5). Of the two spaces $\mathcal{T}_n$ and $\mathcal{P}_n$, the first is easier to handle, for the presence of the endpoints a, b has often a disturbing effect.

The present chapter is almost exclusively devoted to approximation for the space $C(A)$ of continuous real or complex valued functions on A, where the set A is $[a, b]$, or $\mathbb{T}$, or a compact subset of $\mathbb{R}^n$, or, more generally, a compact metric or a compact Hausdorff topological space. The norm on $C(A)$ is defined by

$$\|f\| := \sup_{x \in A} |f(x)| = \max_{x \in A} |f(x)| .$$

Convergence in the space $C(A)$ is convergence with respect to the norm; $C(A)$ is a Banach space. We assume known basic facts of Banach space theory; for details see also Chapter 2.

Our simplest theorems are the two theorems of Weierstrass:

Theorem 1.1 (Weierstrass, 1885). *Each continuous real function f on $[a,b]$ is uniformly approximable by algebraic polynomials: for each $\varepsilon > 0$ there is some P with*

$$|f(x) - P(x)| < \varepsilon, \qquad a \le x \le b.$$

Theorem 1.2. *For each function $f \in C(\mathbb{T})$ and each $\varepsilon > 0$ there is a trigonometric polynomial T for which*

$$|f(x) - T(x)| < \varepsilon, \qquad x \in \mathbb{T}.$$

The proofs of these theorems given in §2 will depend upon properties of some special linear operators. For a function $f \in C[0,1]$, the formula

$$B_n(f,x) := \sum_{k=0}^{n} \binom{n}{k} x^k (1-x)^{n-k} f\left(\frac{k}{n}\right), \qquad n = 0, 1, \dots \tag{1.3}$$

produces a linear map $f \to B_n(f)$ of $C[0,1]$ into $\mathcal{P}_n$. This is the *Bernstein polynomial* of f (Bernstein, 1912). It is a positive (i.e., satisfies $B_n(f) \ge 0$ if $f \ge 0$) bounded operator of norm 1, because for $x \in [0,1]$,

$$\sum_{k=0}^{n} p_{n,k}(x) = 1, \quad p_{n,k}(x) := \binom{n}{k} x^k (1-x)^{n-k} \ge 0. \tag{1.4}$$

It is instructive to compute $B_n(e_k)$, $e_k(x) = x^k$, $k = 0, 1, 2$. We have

$$\begin{aligned} \sum_{k=0}^{n} k\, p_{n,k}(x) &= \sum_{k=1}^{n} k \binom{n}{k} x^k (1-x)^{n-k} \\ &= nx \sum_{j=0}^{n-1} \binom{n-1}{j} x^j (1-x)^{n-1-j} = nx. \end{aligned} \tag{1.5}$$

Also,

$$\begin{aligned} \sum_{k=0}^{n} k(k-1) p_{n,k}(x) &= n(n-1)x^2 \sum_{j=0}^{n-2} \binom{n-2}{j} x^j (1-x)^{n-2-j} \\ &= n(n-1)x^2, \end{aligned}$$

so that

$$\sum_{k=0}^{n} k^2 p_{n,k}(x) = n^2x^2 + nx(1-x). \tag{1.6}$$

From formulas (1.4), (1.5), (1.6) we derive

$$B_n(e_0) = e_0, \quad B_n(e_1) = e_1, \quad B_n(e_2, x) = e_2(x) + \frac{x(1-x)}{n}. \tag{1.7}$$

A corollary of this is that $B_n(l) = l$ for every linear function l.

Let $f \in L_1(\mathbb{T})$ be an integrable function on $\mathbb{T}$. The coefficients of its Fourier series

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx), \tag{1.8}$$

are given by the formulas

$$a_k := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos kt\, dt, \quad b_k := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \sin kt\, dt.$$

For example, the finite sum (1.2), augmented by zero terms for $k > n$ is the Fourier series of the trigonometric polynomial T_n. If s_n is the n-th partial sum of (1.8), we write $s_n = u_0 + u_1 + \cdots + u_n$, where

$$u_0 := \frac{1}{\pi} \int_{-\pi}^{\pi} \frac{1}{2} f(t) dt, \quad u_k := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \cos k(x-t) dt, \quad k = 1, 2, \dots$$

and apply the formula

$$1/2 + \cos\alpha + \cdots + \cos n\alpha = \frac{\sin(2n+1)\alpha/2}{2\sin(\alpha/2)} =: D_n(\alpha), \tag{1.9}$$

which is obtained by multiplying both sides with $\sin(\alpha/2)$. Thus,

$$\begin{aligned} s_n = s_n(f,x) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) \frac{\sin(2n+1)\frac{x-t}{2}}{\sin\frac{x-t}{2}} dt \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) D_n(x-t) dt. \end{aligned} \tag{1.10}$$

In the same way, by means of the formula

$$\sin\frac{\alpha}{2} + \sin\frac{3}{2}\alpha + \cdots + \sin(2n-1)\frac{\alpha}{2} = \frac{\sin^2(n\alpha/2)}{\sin(\alpha/2)}, \tag{1.11}$$

we obtain a representation of the arithmetic means $\sigma_n := (s_0 + \cdots + s_n)/(n+1)$ of the s_n:

$$\begin{aligned} \sigma_n &:= \sigma_n(f,x) := \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) F_n(x-t) dt, \\ F_n(\alpha) &:= \frac{1}{2(n+1)} \frac{\sin^2\frac{(n+1)\alpha}{2}}{\sin^2\frac{\alpha}{2}}. \end{aligned} \tag{1.12}$$

The expressions $D_n(\alpha)$ and $F_n(\alpha)$ are called the Dirichlet and the Fejér kernels, respectively. They are even trigonometric polynomials of degree n. The kernel F_n is positive, while D_n changes sign. Since $\sigma_n(e_0) = e_0$ for $e_0(x) = 1$, we have $\|\sigma_n\| = 1$. In contrast, $\|s_n\| \to \infty$ (see Theorem 2.1 of Chapter 9).

For $t_k(x) := \cos kx$, from the definition of σ_n,

$$\sigma_n(t_k) = \frac{n-k+1}{n+1} t_k\,, \quad n \geq k$$

with the same formula for $t_k^*(x) := \sin kx$. Therefore,

$$\sigma_n(t_k) \to t_k, \quad \sigma_n(t_k^*) \to t_k^*\,, \quad k = 0, 1, \ldots, \text{ for } n \to \infty\,. \tag{1.13}$$

Weierstrass theorems are valid also in many other Banach spaces. Let X be a Banach space of measurable functions on $[a,b]$ with norm $\|\cdot\|_X$. Often $C[a,b]$ is densely imbedded in X. This means that

$$\|g\|_X \leq M\|g\|_{C[a,b]}\,, \quad g \in C[a,b] \tag{1.14}$$

with some constant M depending on X, and that for each $f \in X$ and each $\varepsilon > 0$, there is a $g \in C[a,b]$ with $\|f - g\|_X < \varepsilon$. For instance, the spaces $L_p[a,b]$, $1 \leq p < \infty$ have this property. If $P \in \mathcal{P}_n$ uniformly approximates the continuous function g with error $< \varepsilon$, then

$$\|f - P\|_X \leq \|f - g\|_X + \|g - P\|_X \leq (1+M)\varepsilon\,,$$

and we obtain a Weierstrass theorem for X. Thus, we have Weierstrass theorems for all $L_p[a,b]$, $1 \leq p < \infty$. Since L_∞ is not separable, this cannot be true in this case. Similar remarks hold for trigonometric approximation of 2π periodic functions.

§ 2. Approximation by Integral Operators

Let $A = [a,b]$ or $A = \mathbb{T}$, and let $K_n(x,y)$ be a sequence of continuous functions ("kernels") for $x, y \in A$. Then, the functions

$$f_n(x) := \int_A K_n(x,t) f(t)dt\,, \quad n = 1, 2, \ldots \tag{2.1}$$

are defined on A for each integrable function $f \in L_1(A)$. We can hope that $f_n(x)$ is close to $f(x)$ if the kernel $K_n(x,t)$ is very small for $x \neq t$ and large n and its total mass approximates 1. In this section, we assume that

$$\int_A K_n(x,t)dt \to 1\,, \qquad \text{uniformly in } x \text{ for } n \to \infty\,, \tag{2.2}$$

$$\int_{|x-t|\geq\delta} |K_n(x,t)|\,dt \to 0\,, \qquad \text{uniformly in } x \text{ for } n \to \infty \tag{2.3}$$

for each $\delta > 0$. (For $A = \mathbb{T}$, the integral is over $\delta \leq |x - t| \leq \pi$.) It follows from (2.3) that for each $\delta > 0$ and $f \in C(A)$,

$$\int_{|x-t|\geq\delta} f(t)K_n(x,t)dt \to 0\,, \quad \text{uniformly in } x \text{ for } n \to \infty\,. \tag{2.4}$$

Theorem 2.1. *In addition to* (2.2), (2.3) *let*

$$\int_A |K_n(x,t)|\,dt \leq M(x) < +\infty\,, \text{ for each } x \in A\,, \quad n = 1, 2, \ldots\,. \tag{2.5}$$

Then, (i) *$f_n(x) \to f(x)$ for each continuous f and each $x \in A$;* (ii) *This convergence is uniform if $M(x)$ does not depend on x.*

Proof. Let $\varepsilon > 0$ be given and let $\delta > 0$ be so small that $|f(t) - f(x)| < \varepsilon$ for $|x - t| \leq \delta$. Because of (2.2),

$$f_n(x) - f(x) = \int_A [f(t) - f(x)]K_n(x,t)dt + o(1)\,, \tag{2.6}$$

where the last term $\to 0$ uniformly for $n \to \infty$. We have

$$\left|\int_{|x-t|\leq\delta} [f(t) - f(x)]K_n(x,t)dt\right| \leq \varepsilon \int_{|x-t|\leq\delta} |K_n(x,t)|\,dt$$
$$\leq \varepsilon M(x)\,,$$

and for the integral over $|x - t| \geq \delta$, we use (2.4). Hence

$$|f_n(x) - f(x)| \leq \varepsilon M(x) + o(1)\,,$$

and (i) and (ii) follow. □

For example, if $K_n(x,t) := \frac{1}{\pi} F_n(x - t)$ where F_n is the Fejér kernel, then (2.3) follows from

$$F_n(x-t) \leq \frac{1}{2(n+1)\sin^2\frac{\delta}{2}}\,, \qquad \text{for } \delta \leq |t - x| \leq \pi\,,$$

and (2.2) follows from $\int_{\mathbb{T}} F_n dt = \pi$; this implies also (2.5) with $M(x) = 1$, because $F_n \geq 0$.

Corollary 2.2. *For the Fejér sums σ_n, we have $\sigma_n(f) \to f$ in the norm of $C(\mathbb{T})$ for each continuous f.*

This also proves Theorem 1.2.

There is another proof of this *if Theorem 1.2 is known*. From the discussion at the end of §1, $\sigma_n(T) \to T$ for each trigonometric polynomial T. Since the T are dense in $C(\mathbb{T})$ and the norms of the σ_n are bounded, we have $\sigma_n(f) \to f$ for all f.

There is a companion theorem for operators given by sums, such as (1.3). Without going into details, we prove

Theorem 2.3. *For each $f \in C[0,1]$, we have in the uniform norm,*

$$B_n(f) \to f, \quad n \to \infty. \tag{2.7}$$

Proof. Since $(t-x)^2$ is a linear combination of $e_j(t) = t^j$, $j = 0, 1, 2$, we obtain from (1.7)

$$\begin{aligned}\sum_{\left|\frac{k}{n}-x\right|\geq\delta} p_{n,k}(x) &\leq \frac{1}{\delta^2}\sum_{k=0}^{n}\left(\frac{k}{n}-x\right)^2 p_{n,k}(x)\\ &= \frac{x(1-x)}{n\delta^2} \leq \frac{1}{4n\delta^2}, \quad 0 \leq x \leq 1.\end{aligned} \tag{2.8}$$

This corresponds to (2.3). By the positivity of the $p_{n,k}$, and by (1.4), we have conditions corresponding to (2.2), (2.5). Therefore, the old proof applies. □

This establishes Theorem 1.1 for the interval $[0,1]$, and the case of an arbitrary interval $[a,b]$ follows from this by a linear substitution.

One can go further and prove convergence $\sigma_n(f,x) \to f(x)$ for discontinuous functions f. We say that $x \in A$ is a *Lebesgue point* of a function $f \in L_1(A)$ if

$$\frac{1}{h}\int_x^{x+h} |f(t) - f(x)|dt \to 0, \quad \text{for } h \to 0. \tag{2.9}$$

It is known that almost all points of A are Lebesgue points of a function $f \in L_1(A)$.

A function $L(t) \geq 0$ on $\mathbb{T}$ is *bell-shaped* at $x \in \mathbb{T}$, if it increases on $(x-\pi, x)$, and decreases on $(x, x+\pi)$.

Theorem 2.4. *If the kernel $K_n(x,t)$ on $\mathbb{T}$ satisfies (2.2), and*

$$\max_{\delta\leq|x-t|\leq\pi} |K_n(x,t)| \to 0 \tag{2.10}$$

uniformly in x for each $\delta > 0$, and if K_n has a majorant $L_n(x,t) : |K_n(x,t)| \leq L_n(x,t)$, which is bell shaped at x and satisfies

$$\int_T L_n(x,t)dt \leq M(x) < +\infty \tag{2.11}$$

then $f_n(x) \to f(x)$, at each Lebesgue point x of f.

Remark. This result also holds for $A = [a,b]$ in place of $\mathbb{T}$ with a similar proof.

Proof. We begin with the representation (2.6). From (2.10), we have (2.4) for each $f \in L_1$. For the integral over $(x, x+\delta)$, we apply twice integration by parts. Let

$$F(u) := \int_x^u |f(t) - f(x)|dt, \quad u > x.$$

For $\varepsilon > 0$, we take $\delta > 0$ with the property $|F(u)| < \varepsilon(u-x)$ if $u \in (x, x+\delta]$. Then,

$$\begin{aligned}\left|\int_x^{x+\delta}[f(t)-f(x)]K_n(x,t)dt\right| &\leq \int_x^{x+\delta}|f(t)-f(x)|L_n(x,t)dt\\ &= L_n(x,x+\delta)F(x+\delta) + \int_x^{x+\delta} F(t)\, d[-L_n(x,t)]\\ &\leq \varepsilon\delta L_n(x,x+\delta) + \varepsilon\int_x^{x+\delta}(t-x)d[-L_n(x,t)]\\ &= \varepsilon\delta L_n(x,x+\delta) - \varepsilon\delta L_n(x,x+\delta)\\ &\quad + \varepsilon\int_x^{x+\delta} L_n(x,t)dt \leq \varepsilon M(x).\end{aligned}$$

In a similar fashion, we estimate the integral over $(x-\delta, x)$. The proof is completed as in Theorem 2.1. □

We prove that the Fejér kernel of (1.12) satisfies the conditions of our theorem. Since $\sin^{-2} u \geq u^{-2}$, and ≥ 1, we have

$$\sin^{-2} u \geq \frac{1}{2}(1+u^{-2}) = \frac{u^2+1}{2u^2}.$$

Hence

$$\sin^2\frac{(n+1)(t-x)}{2} \leq \frac{2(n+1)^2(t-x)^2}{(n+1)^2(t-x)^2+4}.$$

Also,

$$\sin^2\frac{t-x}{2} \geq \frac{(t-x)^2}{\pi^2},$$

and we obtain

$$F_n(t-x) \leq \frac{(n+1)\pi^2}{(n+1)^2(t-x)^2+4} =: L_n(x,t).$$

From this (2.10) and (2.11) follow easily.

Corollary 2.5. *For a function $f \in L_1(\mathbb{T})$, one has $\sigma_n(f,x) \to f(x)$ a.e.*

The situation with the convergence almost everywhere of $s_n(f,x)$ to $f(x)$ is infinitely more complex. In 1923, Kolmogorov constructed a function $f \in L_1(\mathbb{T})$, for which $s_n(f,x)$ is everywhere divergent. Only in 1966, did L. Carleson show that $s_n(f,x) \to f(x)$ a.e. for each continuous function $f \in C(\mathbb{T})$, and this proved to be true even for $f \in L_p(\mathbb{T})$, $p > 1$. The proof is highly complicated and does not pinpoint a simple property of the points of convergence.

Finally, we mention an important general method for constructing kernels which satisfy the conditions of Theorem 2.1 Let Λ be in $L_1(\mathbb{R})$, have integral one: $\int \Lambda(u)du = 1$. For $\varepsilon > 0$, the dilated functions $\Lambda_\varepsilon(u) := \varepsilon^{-1}\Lambda(u/\varepsilon)$ also have integral one and for any $\delta > 0$, they satisfy

$$\int_{|x-t|\geq\delta} |\Lambda_\varepsilon(x-t)|\,dt = \int_{|u|\geq\delta/\varepsilon} |\Lambda(u)|du \to 0\,, \quad \varepsilon \to 0\,.$$

Hence the kernels $\Lambda_\varepsilon(x-t)$ satisfy (2.1), (2.2) with the parameter n replaced by ε^{-1}. It follows that for any measurable bounded function f defined on $\mathbb{R}$,

$$f_\varepsilon(x) := \int_{\mathbb{R}} f(t)\Lambda_\varepsilon(x-t)dt \to f(x)\,, \quad \varepsilon \to 0\,, \tag{2.12}$$

at each point of continuity of f and the convergence is uniform on $\mathbb{R}$ if f is uniformly continuous there. If f is continuous on $A = [a,b]$ and we extend the definition of f to all of $\mathbb{R}$ by setting $f(x) := 0$, $x \notin A$, then the convergence in (2.12) holds for all $x \in (a,b)$ and is uniform on any closed subinterval of (a,b).

§ 3. The Theorem of Korovkin

There is another approach to theorems of Weierstrass type. In the space $C(A)$, where A is a compact Hausdorff space, we can distinguish positive elements: $f \geq 0$ means that $f(x) \geq 0$ for all $x \in A$. We write $f \leq g$ for $g - f \geq 0$ and $|f|$ for the function with values $|f(x)|$, $x \in A$. An operator U that maps C into itself is *positive* if $f \geq 0$ implies $U(f) \geq 0$. If in addition, U is linear, we have $U(f) \leq U(g)$ if $f \leq g$; consequently $|U(f)| \leq U(|f|)$ for all $f \in C$. An operator of this type is always bounded: from

$$|U(f)| \leq U(|f|) \leq U\left(\|f\|e_0\right) = \|f\|U\left(e_0\right)$$

(where e_0 is the constant function 1), it follows that $\|U(f)\| \leq \|U(e_0)\| \cdot \|f\|$. This also gives a formula for its norm

$$\|U\| = \|U(e_0)\|\,. \tag{3.1}$$

Korovkin [1957] has shown that for a sequence (U_n) of positive linear operators, in many cases, convergence $U_n(f) \to f$ in the uniform norm follows for all $f \in C(A)$, if it holds for some finitely many "test functions" $g_1, \ldots, g_m$ from $C(A)$. Here is a version of his theorem from Lorentz [A-1966].

Theorem 3.1. *Assume that there exist continuous real functions $a_i(y)$ on A, $i = 1, \ldots, m$, such that*

$$P_y(x) = \sum_{i=1}^m a_i(y)g_i(x) \geq 0 \tag{3.2}$$

for all $x, y \in A$ and that $P_y(x) = 0$ holds if and only if $x = y$. Then for a sequence of positive linear operators U_n on $C(A)$, the convergence

$$U_n(g_i) \to g_i\,, \quad n \to \infty\,, \quad i = 1, \ldots, m \tag{3.3}$$

implies that

$$U_n(f) \to f\,, \quad n \to \infty\,, \tag{3.4}$$

for all $f \in C(A)$.

Proof. We begin with some properties of the functions $P(x) = \sum_{i=1}^m a_i g_i(x)$. There exists a P^* with $P^*(x) > 0$ for all $x \in A$: if $y_1 \neq y_2$ are two points of A, we can take $P^* := P_{y_1} + P_{y_2}$. From (3.3), we have $U_n(P,x) \to P(x)$ uniformly in x for each P with constant coefficients. We also have

$$U_n(P_y, y) = \sum_{i=1}^m a_i(y)U_n(g_i, y) \to \sum_{i=1}^m a_i(y)g_i(y) = 0\,,$$

and the convergence is uniform in y because the $a_i(y)$ are bounded. Finally, for some constant $M_0 > 0$, $\|U_n(e_0)\| \leq M_0$. This follows from

$$U_n(e_0, x) \leq a\,U_n(P^*, x) \to a\,P^*(x)\,,$$

where $a > 0$ is taken so that $1 = e_0(x) \leq a\,P^*(x)$, $x \in A$.

We need the following fact. Let $f_y \in C(A)$, $y \in A$, be a family of functions for which $f_y(x)$ is a continuous function of $(x,y) \in A \times A$ and $f_y(y) = 0$ for all $y \in A$. Then

$$U_n(f_y, y) \to 0\,, \quad \text{uniformly in } y\,, \quad \text{for } n \to \infty\,. \tag{3.5}$$

For the proof, consider the "diagonal" set $B := \{(y,y) : y \in A\}$ in $A \times A$ and some $\varepsilon > 0$. Each point of B has a neighborhood V in $A \times A$ for which $|f_y(x)| < \varepsilon$ if $(x,y) \in V$. The union G of all these V is an open set; its complement F is compact. Let

$$m := \min_{(x,y)\in F} P_y(x) > 0\,, \quad M := \max_{(x,y)\in F} |f_y(x)|\,.$$

For all x, y, we have

$$|f_y(x)| \leq \varepsilon + \frac{M}{m}P_y(x)\,. \tag{3.6}$$

In fact, $|f_y(x)|$ does not exceed the first term on the right if $(x,y) \in G$, nor the second term if $(x,y) \in F$. From (3.6), we derive

$$\begin{aligned}|U_n(f_y,y)| &\le \varepsilon U_n(e_0,y) + \frac{M}{m}U_n(P_y,y)\\ &\le M_0\varepsilon + \frac{M}{m}U_n(P_y,y) \le (M_0+1)\varepsilon\,,\end{aligned}$$

for all large n; this proves (3.5).

Now, the proof of the theorem can be completed easily. If $f \in C(A)$ is given, we put

$$f_y(x) := f(x) - \frac{f(y)}{P^*(y)}P^*(x)\,.$$

By what was shown,

$$U_n(f,y) - \frac{f(y)}{P^*(y)}U_n(P^*,y) \to 0\,,$$

uniformly in y and since $U_n(P^*,y) \to P^*(y)$, uniformly, we obtain (3.4). □

Theorem 3.1 gives an alternate approach to Weierstrass' theorems. If $g_1 = 1$, $g_2 = x$, $g_3 = x^2$ on $[0,1]$, a possible function (3.2) is

$$P_y(x) = (y-x)^2 = y^2g_1 - 2yg_2 + g_3\,.$$

Thus, for the Bernstein operators B_n, Theorem 2.3 follows from the relations (1.7). Similarly, on $\mathbb{T}$, the functions $1, \cos x, \sin x$ are an admissible set of test functions, for we can take $P_y(x) = 1 - \cos(y-x)$. Therefore, from (1.13) for $k = 0, 1$, we obtain Corollary 2.2.

Another example is provided by the m-dimensional Bernstein polynomials, which are positive linear operators defined on the m-dimensional cube $Q := [0,1]^m$:

(3.7)

$$B_n(f,x_1,\dots,x_m) := \sum_{k_1=0}^{n}\cdots\sum_{k_m=0}^{n} f\left(\frac{k_1}{n},\dots,\frac{k_m}{n}\right)p_{n,k_1}(x_1)\dots p_{n,k_m}(x_m)\,.$$

We prove that $B_n(f) \to f$ for each $f \in C(Q)$. Indeed, this is true for each of the $2m+1$ functions 1, x_i, x_i^2, $i = 1,\dots,m$, because of (1.7). For (3.2), we can take

$$P_y(x) := \sum_{i=1}^{m}(y_i - x_i)^2\,;$$

we have $P_y(x) \ge 0$ and $P_y(x) = 0$ if and only if $y = x$ for $y = (y_1,\dots,y_m)$, $x = (x_1,\dots,x_m)$.

From this, we derive the m-dimensional Weierstrass theorem:

Theorem 3.2. *A continuous function $f(x_1,\dots,x_m)$, defined on a compact subset A of $\mathbb{R}^m$ is uniformly approximable by polynomials in $x_1,\dots,x_m$.*

Proof. We can imbed A into a cube Q, extend f by Teitze's theorem to a continuous function on Q, and invoke the statement about Bernstein polynomials. □

§4. Theorems of Stone-Weierstrass

We discuss here another wide generalization of Weierstrass' theorem. We approximate real continuous functions $f \in C(A)$, defined on a compact Hausdorff space A. A linear space $G \subset C(A)$ is a *real algebra* of continuous functions, if it is invariant under multiplication, that is, if $g_1, g_2 \in G$ imply $g_1g_2 \in G$. Examples of algebras are all real polynomials on $[-1,1]$ or all real trigonometric polynomials on $\mathbb{T}$.

For a given G, we want to know whether all $f \in C(A)$ are arbitrarily closely approximable by elements $g \in G$, and if not then which f are? The answers are provided by the Stone-Weierstrass theorems that follow.

Proofs of Stone-Weierstrass theorems depend on some elementary inequality or approximation theorem, and on topological tools. In the present proof (based on Brosowski-Deutsch [1981] and Ransford [1984]), the inequalities used will be

$$(4.1)\quad 1 - Nu \le (1-u)^N \le \frac{1}{(1+u)^N} \le \frac{1}{Nu}\,,\quad 0 < u \le 1\,,\quad N = 1,2,\dots\,.$$

The topological tool will be the distance $d_B(f)$ from $f \in C(A)$ to G on a non-empty compact subset B of A:

$$(4.2)\qquad d_B(f) := \inf_{g\in G}\|f - g\|_{C(B)}\,.$$

A compact set $B \subset A$ will be called a G-constancy set if each $g \in G$ is constant on B. A useful remark is that *if a compact set $B \subset A$ is not a G-constancy set, then there is an $h \in G$ for which*

$$(4.3)\qquad 0 \le \mu := \min_{x\in B} h(x) < \max_{x\in B} h(x) = 1\,.$$

Indeed, let $h_1 \in G$ be non-constant on B. We can assume that $\|h_1\| = 1$. If h_1^2 is not constant on B, we take $h := h_1^2$. Otherwise h_1 takes only the values ± 1 on B, and we put $h := \frac{3}{4}h_1^2 + \frac{1}{4}h_1$.

Lemma 4.1 (Machado [1959]). *For each $f \in C(A)$, there is a minimal non-empty compact set $A_0 \subset A$ for which*

$$(4.4)\qquad d_{A_0}(f) = d_A(f)\,;$$

the set A_0 is a G-constancy set.

Proof. If $d_A(f) = 0$, then each one point set is the set A_0. We assume therefore that $d_A(f) > 0$. Let F be the family of all nonempty compact sets $B \subset A$ such that $d_B(f) = d_A(f)$. It contains the set A. If F_0 is a subfamily of F ordered by inclusion, then $B_0 := \cap_{B \in F_0} B$ also belongs to F. Indeed, for any g and any $B \in F_0$, $\{x \in B : |f(x) - g(x)| \geq d_A(f)\}$ is compact and non-empty. The intersection of these sets, namely $\{x \in B_0 : |f(x) - g(x)| \geq d_A(f)\}$ also has these properties, so that $d_{B_0}(f) = d_A(f)$. By Zorn's lemma therefore, there is a minimal element A_0 in F.

We have to show that A_0 is a G-constancy set. Suppose this is not the case. Then, there exists a function $h \in G$ satisfying (4.3). Putting $\mu := 1 - 3\delta$, $0 < \delta \leq 1/3$, we consider the two non-empty compact sets

$$B_1 := \{x \in A_0 : 1 - 3\delta \leq h(x) \leq 1 - \delta\} ,$$
$$B_2 := \{x \in A_0 : 1 - 2\delta \leq h(x) \leq 1\} .$$

Let $B_0 := B_1 \cap B_2$. On the sets $B_1 \backslash B_2$, B_0, $B_2 \backslash B_1$, we have respectively

$$1 - 3\delta \leq h(x) < 1 - 2\delta ; \quad 1 - 2\delta \leq h(x) \leq 1 - \delta ; \quad 1 - \delta < h(x) \leq 1 .$$

Each of the sets B_i, $i = 1, 2$ is a proper subset of A_0. Therefore, there are functions g_i, $i = 1, 2$ in G with

$$|f(x) - g_i(x)| < d_A(f) , \quad i = 1, 2 , \quad x \in B_i . \tag{4.5}$$

Now, let for $n = 3, 4, \ldots$,

$$g_n := H_n g_1 + (1 - H_n) g_2 , \quad H_n := (1 - (h(x))^n)^{N_n} .$$

It is possible that $H_n \notin G$, for we have not assumed that G contains the functions 1, but $g_n \in G$ in any case. The integers N_n we select so that $N_n(1 - 2\delta)^n \to 0$, $N_n(1 - \delta)^n \to \infty$.

On B_0, we have both inequalities (4.5), and since g_n is a convex combination of g_1 and g_2,

$$|f(x) - g_n(x)| < d_A(f) , \quad n = 3, 4, \ldots . \tag{4.6}$$

We use inequalities (4.1): for $x \in B_1 \backslash B_2$,

$$H_n(x) \geq 1 - N_n(h(x))^n \geq 1 - N_n(1 - 2\delta)^n \to 1 ,$$

and for large n, g_n is arbitrarily close to g_1. For $x \in B_2 \backslash B_1$,

$$H_n(x) \leq \frac{1}{N_n(h(x))^n} \leq \frac{1}{N_n(1 - \delta)^n} \to 0$$

and g_n is arbitrarily close to g_2. This yields $|f(x) - g_n(x)| < d_A(f)$, provided n is sufficiently large. This is a contradiction. □

By means of the lemma, we obtain

Theorem 4.2. *A function $f \in C(A)$ is arbitrarily closely approximable on A by functions $g \in G$ if and only if on each constancy set B, it coincides with some $g \in G$.*

Proof. If f is approximable on A, then it is approximable on B, so that $f = C$ is a constant on B. If $C = 0$, the assertion is clear. If $C \neq 0$, then there must also be some $g \neq 0$, and then one of its multiples equals f on B. Conversely, the assumption implies that f is approximable on the set A_0 of Lemma 4.1. Then $d_{A_0}(f) = 0$, hence $d_A(f) = 0$. □

From this, the usual form of the Stone-Weierstrass theorem follows at once. We say that a function f (or an algebra G) separates two points x, y of A if $f(x) \neq f(y)$ (or if $g(x) \neq g(y)$ for some $g \in G$).

Theorem 4.3 (Stone [1948]). *A function $f \in C(A)$ is approximable by functions $g \in G$ of a real algebra G if and only if:*

(a) *$g(x) = 0$ for all $g \in G$ implies $f(x) = 0$, and*
(b) *The algebra G separates any two points separated by f.*

Proof. The necessity of the two conditions is obvious. It remains to show that they imply the conditions of Theorem 4.2. Let B be a G-constancy set. By (b), f is constant on B. We need to consider only the case when f does not vanish on this set. By (a), some $g \in G$ also has this property. Then, f is a constant multiple of this g. □

Here are some applications.

Example 1. The even trigonometric polynomials on $[0, \pi]$ form an algebra containing the function 1 and the function $\cos x$ which separates points of $[0, \pi]$. These polynomials are dense in $C[0, \pi]$. On the other hand, they are not dense in $C[0, a]$, for $a > \pi$.

Example 2. The algebra of all trigonometric polynomials contains 1 and the two functions $\cos x$, $\sin x$, which together separate points of $\mathbb{T}$. Therefore, it is dense in $C(\mathbb{T})$.

Example 3. The algebraic polynomials are dense in $C[0, 1]$ and in $C[0, 1]^m$ because the function x separates points in $[0, 1]$, and $x_1, \ldots, x_m$ separate points in $[0, 1]^m$.

Example 4. Let Π be the topological product $\Pi := \Pi_{j \in J} A_j$ of some compact Hausdorff spaces A_j. The functions $\phi = f_{j_1} \ldots f_{j_m}$ with $f_{j_k} \in C(A_{j_k})$ separate points, but they do not form an algebra. Sums of products $g = \sum_{k=1}^{n} \phi_k$ do; therefore they approximate all $f \in C(\Pi)$.

For complex valued functions, Theorem 4.3 is false. For example, the polynomials with complex coefficients $P(z) = \sum_{k=0}^{n} a_k z^k$ form an algebra G on the disk $|z| \leq 1$ which satisfies the conditions of Theorem 4.3, but the uniform closure of G consists only of analytic functions.

There is, however, a formulation of our theorems which remains valid in the complex case. In the space $C(A)$ of complex valued functions on A, let $G \subset C(A)$ be a complex algebra. A set $B \subset A$ is called *antisymmetric* (with respect to G) if each function $g \in G$ that is real on B is also constant on B. For example, if $A \subset C$ and G is an algebra of analytic functions, then each open connected subset of A is antisymmetric.

Examining the proof of Lemma 4.1, we see that it remains valid if constancy sets are replaced by antisymmetric sets. The proof of Theorem 4.2 then gives

Theorem 4.4 (Bishop [1961]). *A complex valued function $f \in C(A)$ is uniformly approximable by the elements of G if and only if it is uniformly approximable on each antisymmetric set of A; if G is a closed algebra then $f \in C(A)$ is uniformly approximable by the elements of G if and only if f coincides with a function $g \in G$ on each antisymmetric set.*

§ 5. Problems

5.1. A continuous function $f(x)$, $-\infty < x < +\infty$, which has a finite limit at infinity, $\lim_{x\to\pm\infty} f(x)$, can be approximated on $(-\infty, +\infty)$ by rational functions of the form $(1+x^2)^{-n}P_{2n}(x)$.

5.2. Let Q_n be a sequence of polynomials of degrees m_n, and $Q_n(x) \to f(x)$ uniformly on $[a,b]$. If f is not a polynomial, then $m_n \to \infty$.

5.3. Under the same assumptions as in Problem 5.2, $|E_n| \to 0$, where E_n is the set of points x for which $Q_n(x) = f(x)$. (Brudnyi and Gopengauz)

5.4. If $f \in C[0,1]$ and $f(0) = f(1) = 0$, then the sequence of polynomials with integral coefficients

$$\sum_{k=1}^{n-1} \left[\binom{n}{k} f\left(\frac{k}{n}\right) \right] x^k (1-x)^{n-k}$$

converges uniformly to $f(x)$.

5.5. A function $f \in C[0,1]$ is approximable by polynomials with integral coefficients if and only if $f(0)$ and $f(1)$ are integers.

5.6. If the function f is continuous on $[0,\infty)$ and has limit zero for $x \to +\infty$, then $\lim_{u\to\infty} S_u(x) = f(x)$ uniformly for $0 \le x < +\infty$, where $S_u(x)$ is defined by

$$S_u(x) := e^{-ux} \sum_{k=0}^{\infty} \frac{(ux)^k}{k!} f\left(\frac{k}{u}\right), \quad u > 0 .$$

5.7. If in Theorem 2.4 the kernel K_n itself is bell-shaped, then $f_n(x) \to f(x)$, $n \to \infty$ at each x where f is the derivative of its indefinite integral. In particular, this applies to the Poisson transform of f for $r \to 1$, which has the kernel

$$K_r(x,t) = \frac{1}{2\pi} \frac{1-r^2}{1-2r\cos(x-t)+r^2}, \quad 0 \le r < 1 .$$

5.8. If f is continuous on $\left[-\frac{1}{2}, \frac{1}{2}\right]$, let $f_n(x) := \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t)\lambda_n(t-x)dt$ with $\lambda_n(t) := c_n\left(1-t^2\right)^n$ and c_n chosen so that $\int_{-1}^{1} \lambda_n(t)dt = 1$. Then f_n is an algebraic polynomial of degree $\le 2n$ which converges uniformly to f on $\left[-\frac{1}{2}+\delta, \frac{1}{2}-\delta\right]$ for each $0 < \delta < \frac{1}{2}$ (Landau).

§ 6. Notes

6.1. Another example of an integral operator is the Weierstrass integral

$$W_a(f,x) := \frac{1}{\sqrt{\pi}a} \int_{-\infty}^{\infty} e^{-\frac{(x-t)^2}{a^2}} f(t)dt , \quad a > 0 . \tag{6.1}$$

We can reconstruct the original proof (1885) of Weierstrass of Theorem 1.1. We extend $f \in C[0,1]$ to a continuous function $f \in C(\mathbb{R})$ with compact support. From (2.12), it follows that for any given $\varepsilon > 0$ and sufficiently small a, $|f(x) - W_a(f,x)| < \varepsilon$, $x \in [0,1]$. Now, $W_a(f,x)$ is an entire function of x, it is approximable by the partial sums of its Taylor series. This gives a uniform approximation to f on $[0,1]$.

6.2. The following are the two basic approximation theorems for $\mathbb{C}$ which replace the Weierstrass theorems on $\mathbb{R}$. We say that f is analytic on A if it is analytic on some open set containing A. The following theorem of Runge was proved in 1885:

Theorem 6.1. *If A has a connected complement, then each analytic function on A can be approximated in $C(A)$ by polynomials with complex coefficients.*

It is easy to see that the assumption of the connectedness of the complement cannot be omitted. The necessary and sufficient conditions for approximability in the following theorem of Mergelian, 1951, are natural, but the proof is decidedly more difficult.

Theorem 6.2. *If A has a connected complement, then a function $f \in C(A)$ is uniformly approximable by complex polynomials if and only if it is analytic on A^0.*

For the proofs see Gaier [A-1980]. There also exist algebraic-theoretic proofs of Theorem 6.2 (Carleson [1964]).

Chapter 2. Spaces of Functions

§ 1. Introduction. The Spaces C and L_p

The development of Analysis in the beginning of the 20-th century, in particular the study of topological, metric, normed spaces by Fréchet, Hausdorff, F. Riesz, Hilbert, Banach, provided invaluable help for Approximation Theory. It furnished the spaces where the approximation takes place, provided the means to measure the error of approximation, and led to operators which can be used for approximation. In this chapter we treat different spaces not so much for their intrinsic merit, but mainly because of the role that they will play in the rest of the book.

In §§1–2 we shall recapitulate, mostly without proofs, the basic facts of functional analysis. As references we recommend the books of Royden [B-1968], Dunford and Schwartz [B-1958], Yosida [B-1965], Halmos [B-1950] and, for rearrangement-invariant spaces and interpolation theorems, of Bennett and Sharpley [B-1988].

In most cases, our spaces are linear real (or complex) normed function spaces X. The norm $\|\cdot\|$ is a non-negative function on X with the properties

$$(1.1)\qquad \begin{cases} \text{(i)} & \|x\| = 0 \text{ if and only if } x = 0\,, \\ \text{(ii)} & \|ax\| = |a|\,\|x\| \quad \text{for a scalar } a\,, \\ \text{(iii)} & \|x+y\| \le \|x\| + \|y\|\,. \end{cases}$$

If X is complete, it is called a Banach space.

A useful remark is that the space X is complete exactly when each series $\sum_1^\infty x_n$ of elements of X converges in X if $\sum_1^\infty \|x_n\| < \infty$.

A *semi-norm* $|\cdot|$ on X satisfies the conditions (1.1) except that (i) may be violated. The linear subspace $X_0 \subset X$ where $|x| = 0$, is the *null-space* of the semi-norm. For example, in the space $C^r[a,b]$, $r = 1,2,\ldots$, of r-times continuously differentiable functions on $[a,b]$, $|f| = \max_{a\le x\le b} |f^{(r)}(x)|$ is a semi-norm whose null-space is the set $\mathcal{P}_{r-1}$ of all polynomials of degree $\le r-1$. A *quasi-norm* $\|\cdot\|$ on a linear space X also satisfies (1.1), except that (iii) is replaced by

$$(\text{iii}')\qquad \|x+y\| \le C(\|x\| + \|y\|)$$

for some constant $C = C(X)$.

We shall have also quasi-seminorms, with an inequality $|x+y| \le C(|x| + |y|)$. Almost always, our spaces will consist of functions defined on a set A which is a subinterval of $\mathbb{R} = (-\infty,\infty)$ or it will be the unit circle $\mathbb{T}$. Thus, A will be one of the four sets, $\mathbb{R}$, $\mathbb{R}_+ = [0,\infty)$, an interval $[a,b]$ or $\mathbb{T}$. Points of $\mathbb{T}$ are real numbers modulo 2π, but sometimes we identify them with the complex numbers $e^{i\theta}$, $\theta \in \mathbb{R}$.

The real (or complex) space $C(A)$ consists of all real (or complex) valued continuous functions with the norm $\|f\|_\infty = \sup_{x\in A} |f(x)|$. By $\widetilde{C}(A)$ we denote, for $A = \mathbb{R}$ or $\mathbb{R}_+$, the subspace of $C(A)$ of all uniformly continuous functions. If A is compact then

$$(1.2)\qquad \|f\|_\infty := \|f\| := \max_{x\in A} |f(x)|\,.$$

The continuous linear functionals on $C(A)$ are the integrals $l(f) = \int_A f(x)\,d\mu(x)$ for some regular Borel measure on A with $\|l\| = \int |d\mu|$. (See Halmos [B-1950].) For our standard sets $A = \mathbb{R}, \mathbb{R}_+$, $[a,b]$ or $\mathbb{T}$, $C^r(A)$, $r = 1,2,\ldots$ is the space of all r times continuously differentiable functions with the seminorm $\|f^{(r)}\|_\infty$. A norm on C^r is $\|f\| := \|f\|_\infty + \|f^{(r)}\|_\infty$. Finally, $C^\infty(A)$ is the space (which is not normable) of all infinitely differentiable functions on A.

A function g on A is *of bounded variation*, or belongs to the space $BV(A)$ if

$$(1.3)\qquad \operatorname{Var}_A g := \sup \sum_{i=1}^{n-1} |g(x_{i+1}) - g(x_i)| < \infty\,;$$

the supremum is taken for all finite sequences $x_1 < x_2 < \cdots < x_n$ in A. A function g of bounded variation has a representation $g = g_1 - g_2$ with increasing g_1, g_2 which satisfy $\operatorname{Var} g = \operatorname{Var} g_1 + \operatorname{Var} g_2$.

The space BV is important as the dual of the space of continuous functions. Each continuous linear functional $l(f)$, $f \in C(A)$, $A = [a,b]$ or $\mathbb{T}$, is the Stieltjes integral

$$l(f) = \int_A f(x)\,dg(x)$$

for some $g \in BV(A)$. (This is the original F. Riesz' representation theorem.) The norm $\|l\|$ of l is not quite $\operatorname{Var} g$.

Functions $g \in BV$ have at most a countable number of discontinuities and at each point c the left and the right limits $g(c-)$ and $g(c+)$ exist. For some applications, we need functions obtained by changing g at some points c. Firstly, we correct g at all *interior* points $c \in A$ of discontinuity, so that $\bar g(c)$ is between $g(c-)$ and $g(c+)$. The resulting function $\bar g$ yields the norm of the functional l : $\|l\| = \operatorname{Var} \bar g$. Secondly, for the purpose of §9, we need another correction g_0, which is different from $\bar g$ only if $A = [a,b]$ and then only at the endpoints a, b: g_0 must be continuous there. We put $\operatorname{Var}^* g := \operatorname{Var} g_0$.

A function f is *convex* on an interval A in $\mathbb{R}$ if for any two points $x_1, x_2 \in A$ and $x = \alpha x_1 + (1-\alpha)x_2$, $0 \le \alpha \le 1$, we have

$$f(x) \le \alpha f(x_1) + (1-\alpha) f(x_2) . \tag{1.4}$$

Convex functions are continuous at all interior points of A and differentiable except at at most countably many points; at these points, they have one-sided derivatives. A *concave function* satisfies (1.4) with $\le$ replaced by $\ge$.

The sets $A := \mathbb{R}$ or $\mathbb{T}$, are groups under addition. They carry the Lebesgue measure dx, which is the Haar measure with respect to the group operations. Thus, the measure of a set $E \subset A$, $mE := |E|$, $E \subset A$ satisfies: $|E| = |E+a|$, $|aE| = |a|\,|E|$. Similarly, $\mathbb{R}_+ = (0,\infty)$ is a group under multiplication, its Haar measure is $d\mu = \frac{dx}{x}$, with $\mu(E) = \mu(aE)$ for each $a > 0$. On the interval $[-1,1]$, a natural measure is $d\mu = \frac{dx}{\sqrt{1-x^2}}$; this measure is deduced from dt on the upper half, $0 \le t \le \pi$, of $\mathbb{T}$ by the 1–1 map $t \to x$ given by $x = \cos t$ (which maps even trigonometric polynomials into algebraic polynomials). This is one of the explanations of the importance of the function $\sqrt{1-x^2}$ and the measure $d\mu$ for the approximation on the interval $[-1,1]$.

Lebesgue measure on A (which is a subinterval of $\mathbb{R}$ or $\mathbb{T}$) produces the Banach space $L_1(A)$ of integrable functions, by means of $\|f\|_1 := \int_A |f(x)|\,dx$. For example, the *characteristic function* χ_E of a set $E \subset A$ is integrable if $\int_A \chi_E\,dx = |E| < \infty$. If $f \notin L_1(A)$, then possibly f is *locally integrable*, $f \in L_{\mathrm{loc}}(A)$. This means that each point $c \in A$ has a neighborhood on which f is integrable. For $0 < p \le \infty$, the space $L_p(A)$ consists of all measurable functions f for which the following is finite

$$\|f\|_p := \|f\|_p(A) = \begin{cases} \left\{ \int_A |f|^p\,dx \right\}^{1/p}, & 0 < p < \infty \\ \operatorname*{ess\,sup}_{x\in A} |f(x)|, & p = \infty . \end{cases} \tag{1.5}$$

This is a norm for $1 \le p \le \infty$, and $L_p(A)$ is a Banach space. Characteristic for L_p are the inequalities of Hölder and Minkowski. The integral variant of the latter, for measurable functions $g(y) \ge 0$, $y \in B$, $f(x,y) \ge 0$ on $A \times B$, reads

$$\left\{ \int_A \left(\int_B g(y) f(x,y)\,dy \right)^p dx \right\}^{1/p} \le \int_B g(y) \left\{ \int_A f(x,y)^p\,dx \right\}^{1/p} dy . \tag{1.6}$$

The dual space of $L_p(A)$, $1 \le p < \infty$, is the space $L_{p'}(A)$ with the conjugate exponent p', $\frac{1}{p} + \frac{1}{p'} = 1$. Thus, the spaces L_p, $1 < p < \infty$ are reflexive. The spaces L_p with $0 < p < 1$, have a different structure. Trivial examples show that for $p < 1$, (1.5) is not a norm. In this case, the expression $|||f|||_p := \|f\|_p^p = \int_A |f|^p\,dx$ is often useful. Since $|||f+g|||_p = \int_A |f+g|^p\,dx \le |||f|||_p + |||g|||_p$, (1.5) is a quasi-norm with $C = 2^{1/p}$:

$$\begin{aligned} \|f+g\|_p &\le (|||f|||_p + |||g|||_p)^{1/p} \le \left\{ 2\max(|||f|||_p, |||g|||_p) \right\}^{1/p} \\ &\le 2^{1/p}(\|f\|_p + \|g\|_p) . \end{aligned}$$

The discrete analogues of the spaces L_p are the spaces l_p of infinite sequences $\mathbf{x} = (x_i)_{i=1}^\infty$ with the norms

$$\|\mathbf{x}\|_p := \begin{cases} \left(\sum_{i=1}^\infty |x_i|^p \right)^{1/p}, & 0 < p < \infty \\ \sup_{i=1,2,\dots} |x_i| , & p = \infty , \end{cases} \tag{1.8}$$

and their finite dimensional variants $l_p^n \subset \mathbb{R}^n$, with points $\mathbf{x} = (x_i)_{i=1}^n$.

For two positive sequences x_n, y_n we write

$$x_n \approx y , \qquad x_n \sim y_n \tag{1.9}$$

for the *strong*, respectively the *weak equivalence* of x_n and y_n. The first relation means that $x_n/y_n \to 1$, for $n \to \infty$; the second signifies the existence of constants $0 < C_1, C_2 < \infty$ for which $C_1 \le x_n/y_n \le C_2$ for all n. We write $x_n = \mathcal{O}(y_n)$ if $|x_n| \le C y_n$ and $x_n = o(y_n)$ if $x_n/y_n \to 0$. These notations are used also for other ranges of variables.

If $|A| < \infty$, we have the *continuous embedding* of the L_p spaces:

$$L_q(A) \subset L_p(A) , \quad \|f\|_p \le C\|f\|_q , \qquad p \le q .$$

This follows from Hölder's inequality, with $C = |A|^{\frac{1}{p}-\frac{1}{q}}$. For the spaces l_p, we use the counterpart of Hölder's inequality;

$$\left(\sum |x_i|^q \right)^{1/q} \le \left(\sum |x_i|^p \right)^{1/p}, \qquad p \le q , \tag{1.10}$$

which implies the continuous embedding $l_p \subset l_q$.

The spaces L_p, $1 \le p \le \infty$ belong to the category of *Banach function spaces* (or *lattices*). These are Banach spaces X of locally integrable functions f on A, for which the norm $\|\cdot\|_X$ is related to the order by the property: *If* $|f(x)| \le |g(x)|$, *a.e. for* $x \in A$ *and* $g \in X$, *then also* $f \in X$ *and* $\|f\|_X \le \|g\|_X$. One also assumes that the space X is sufficiently rich: it must contain the characteristic functions of all subsets of A of finite measure. The last requirement is the more technical Fatou property: if $f_n \ge 0$ is an increasing sequence of functions in X, and $f_n \to f$ a.e., then $\|f\| = \lim_{n\to\infty} \|f_n\|$. Typical examples of Banach function spaces are L_p spaces with weight $w \ge 0$, with

$$\|f\| := \left(\int_A w(x)|f(x)|^p\,dx \right)^{1/p}, \qquad 1 \le p \le \infty .$$

If X, Y are two Banach function spaces over the same set A with Lebesgue measure, *then* $X \subset Y$ *implies that* X *is continuously embedded in* Y. In other words, for some constant C, one has $\|f\|_Y \le C\|f\|_X$ for all $f \in X$ (see Bennett and Sharpley, [B-1988, p. 7].

For the spaces $L_p(A)$, $0 < p < 1$, equivalent to $\|f\|_p$ is the quasinorm $\|f\|_p^p$, which is subadditive: $\|f+g\|_p^p \le \|f\|_p^p + \|g\|_p^p$. This exemplifies the following general theorem (to be used in Chapters 6, 7, 12).

Theorem 1.1. *For each quasinorm $\|\cdot\|$ on a linear space X there exists an equivalent quasinorm $\|\cdot\|_0$ and a number $\mu > 0$ so that $\|\cdot\|_0^\mu$ is subadditive.*

Proof. With the constant C of (iii′), we take $C_0 := 2C$; we have $C_0 \geq 2$. Then for $f, g \in X$,

$$\|f+g\| \leq C_0 \max(\|f\|, \|g\|) . \tag{1.11}$$

By induction, we derive

$$\|f_1 + \cdots + f_m\| \leq \max_{1\leq j\leq m} (C_0^j \|f_j\|) . \tag{1.12}$$

We now define μ by $C_0^\mu = 2$ and $\|\cdot\|_0$ by

$$\|f\|_0 = \inf_{f=f_1+\cdots+f_m} \{\|f_1\|^\mu + \cdots + \|f_m\|^\mu\}^{1/\mu} , \tag{1.13}$$

where the infimum is taken for all decompositions of f. Clearly, $\|\cdot\|_0^\mu$ is subaddditive, and $\|f\|_0 \leq \|f\|$, $f \in X$. To prove the opposite inequality

$$\|f\| \leq A\|f\|_0 , \qquad f \in X , \tag{1.14}$$

we define:

$$N(f) = \begin{cases} 0 & \text{if } f = 0 \\ C_0^k & \text{if } C_0^{k-1} < \|f\| \leq C_0^k , \quad k \in \mathbb{Z} . \end{cases} \tag{1.15}$$

Since $C_0^{-1} N(f) \leq \|f\| \leq N(f)$, (1.14) will follow from the inequality

$$\|f_1 + \cdots + f_m\| \leq C_0 \{N(f_1)^\mu + \cdots + N(f_m)^\mu\}^{1/\mu} . \tag{1.16}$$

The proof of (1.16) is by induction. This is immediate if $m = 1$. We suppose that (1.16) has been established for $m = n - 1$ and consider any $f_1, \ldots, f_n \in X$. We can assume that $\|f_1\| \geq \cdots \geq \|f_n\|$. If all of the values $N(f_j)$, $j = 1, \ldots, n$, are distinct, we have

$$C_0^j \|f_j\| \leq C_0^j N(f_j) \leq C_0 N(f_1) \leq C_0 \{N(f_1)^\mu + \cdots + N(f_n)^\mu\}^{1/\mu}$$

and therefore (1.16) follows from (1.12). On the other hand, if $N(f_j) = N(f_{j+1}) = C_0^l$ for some $1 \leq j < n$, and $l \in \mathbb{Z}$, then by (1.11), $\|f_j + f_{j+1}\| \leq C_0^{l+1}$. Therefore $N(f_j + f_{j+1})^\mu \leq C_0^{\mu(l+1)} = 2^{(l+1)} = N(f_j)^\mu + N(f_{j+1})^\mu$. Hence, by our induction hypothesis,

$$\begin{aligned} \|f_1 + \cdots + f_n\| &\leq C_0 \{N(f_1)^\mu + \cdots + N(f_j + f_{j+1})^\mu + \cdots + N(f_n)^\mu\}^{1/\mu} \\ &\leq C_0 \{N(f_1)^\mu + \cdots + N(f_j)^\mu + N(f_{j+1})^\mu + \cdots + N(f_n)^\mu\}^{1/\mu} . \end{aligned}$$

The equivalence of $\|\cdot\|$ and $\|\cdot\|_0$ establishes that the second quantity, which is homogeneous, is also a quasinorm. □

§ 2. Rearrangement-Invariant Function Spaces

These are special Banach function spaces; the simplest examples are the L_p spaces. It was Hardy and Littlewood who recognized the importance of the notions of the rearrangement f^* of f and the quasi-order $f \prec g$ (see (2.8) that follows) in describing the norms and operators on these spaces. Rearrangement invariant spaces are important in approximation theory for several reasons: (a) The notions f^* and $f \prec g$ appear in many estimates; (b) Many spaces used for special purposes – such as the Lorentz spaces $L_{p,q}$ – are rearrangement invariant; (c) Several basic approximation theorems, traditionally formulated for C and L_p spaces, are actually valid for arbitrary rearrangement invariant spaces; (d) A more technical reason: the use of the Riesz-Thorin Theorem 4.3 can usually be replaced by the Hardy-Littlewood-Pólya Theorem 4.4 (or Theorem 4.5), which has a simple real proof.

In this section, we shall give an exposition (generally without proofs) of the theory of rearrangements. The proofs are easy and could be supplied by the reader, or found in the book of Bennett and Sharpley [B-1988]. In our development we shall restrict ourselves to the special measure spaces A, $d\mu$, where A is $\mathbb{T}$ or a subinterval of $\mathbb{R}$ and $d\mu$ is Lebesgue measure which will be sufficient for most of our purposes. The other important case where $A = \mathbb{Z}$ and $d\mu$ is counting measured is simpler and can be treated in an analogous fashion. The results of this section actually hold for more general measure spaces provided there is some control on the atoms of the space (see Bennett and Sharpley [B-1988]).

We consider measurable functions f defined on the set $A \subset \mathbb{R}$, equipped with the Lebesgue measure $|E|$ for $E \subset A$. For each f, we define the *distribution function* $\mu_f(y) = \mu_{|f|}(y)$ by

$$\mu(y) := \mu_f(y) := \big|\{x \in A : |f(x)| > y\}\big| , \qquad y \geq 0 . \tag{2.1}$$

The definition would be the same for functions f on any measure space; thus the domain of definition of μ is always $\mathbb{R}_+$. If the function f is bounded and $M :=$ ess sup $|f(x)|$, then $\mu(y) = 0$ outside of $[0, M]$. Two functions $f_1, f_2 \geq 0$ on A are called *equimeasurable*, $f_1 \sim f_2$ if they have the same functions μ. We have:

1. The functions $\mu_f(y)$ have the following properties: (i) *they are non-negative monotone decreasing and right-continuous,* (ii) $|g| \leq |f|$ *a.e. implies* $\mu_g \leq \mu_f$; (iii) *for an increasing sequence* $|f_n| \to |f|$ *one has* $\mu_{f_n} \to \mu_f$.

We prove, for example, the last statement. For a sequence of sets $E_n \subset A$ one defines $\liminf_{n\to\infty} E_n := \bigcup_m \bigcap_{n\geq m} E_m$; then $|\liminf E_n| \leq \liminf_{n\to\infty} |E_n|$. For given $y \geq 0$ we let $E_n := \{x : |f_n(x)| > y\}$ and derive $\mu_{|f|}(y) \leq \liminf \mu_{|f_n|}(y)$. To get (iii), we note that the sequence $\mu_{|f_n|}(y)$ converges and that by (ii) its limit is $\leq \mu_{|f|}(y)$.

For each f on A we define its *decreasing rearrangement* f^* (which is a function on $\mathbb{R}_+$) by

$$f^*(t) = \inf\{y : \mu_f(y) \leq t\} , \qquad t \geq 0 . \tag{2.2}$$

(If the inverse of a monotone, possibly discontinuous function is properly defined, this is exactly the inverse $\mu^{-1}(t)$.) For a simple function f with values c_j, $j = 1, \dots, k$, on disjoint sets, given on a finite interval $[a, b]$, one obtains f^* by replacing the c_j by $|c_j|$ and rearranging the supports to obtain a decreasing function on $[0, b-a]$. Less intuitive are functions on infinite intervals: if $f(x) = \arctan x$, $x \in \mathbb{R}^+$, then $f^*(t) = \frac{\pi}{2}$, $t \geq 0$.

Here are the simplest properties of decreasing rearrangements:

2. (i) $f \sim f^*$; (ii) $|g| \leq |f|$ *a.e. implies* $g^* \leq f^*$; (iii) $(af)^* = |a| f^*$; (iv) $(|f|^p)^* = f^{*p}$, $0 < p < \infty$; (v) *one has* $f_n^* \to f^*$ *if the sequence* $|f_n|$ *is increasing and converges to* $|f|$; (vi) $\int_A |f|\, dx = \int_0^\infty f^*\, dt = \int_0^\infty \mu_f(y)\, dy$; (vii) *weak form of subadditivity:* $(f+g)^*(t_1 + t_2) \leq f^*(t_1) + g^*(t_2)$.

Theorem 2.1. *For two functions* f, g *on* A,

$$\int_A |f(x)g(x)|\, dx \leq \int_0^\infty f^*(t) g^*(t)\, dt\,; \tag{2.3}$$

$$\sup_{\tilde{g} \sim |g|} \int_A |f(x)| \tilde{g}(x)\, dx = \int_0^\infty f^*(t) g^*(t)\, dt\,. \tag{2.4}$$

The supremum here is for all $\tilde{g}$, *equimeasurable with* $|g|$. *For a compact* A, *the supremum in* (2.4) *is achieved. In particular, for each* c, $0 < c < \infty$, *there exists a set* $E_c \subset A$ *with* $|E_c| = c$ *for which*

$$\int_0^c f^*(t)\, dt = \int_{E_c} |f(x)|\, dx\,. \tag{2.5}$$

To prove (2.3), for instance, one first establishes this for simple functions, and then takes a limit, using **2**(v).

If A is not compact, then the local integrability of f does not imply the property (2.5) of f^*. We therefore define the space $L_1(A) + L_\infty(A)$ to consist of all f of the form $f = f_1 + f_2$, $f_1 \in L_1(A)$, $f_2 \in L_\infty(A)$. For compact A, this is simply $L_1(A)$. The above properties yield:

3. *The function* f^* *is locally integrable if and only if* $f \in L_1(A) + L_\infty(A)$.

Often one replaces f^* by the function f^{**}, which is defined by

$$f^{**}(t) = \frac{1}{t} \int_0^t f^*(u)\, du \qquad t > 0\,. \tag{2.6}$$

The main property of f^{**} (lacking in f^*), is its subadditivity:

$$(f_1 + f_2)^{**}(t) \leq f_1^{**}(t) + f_2^{**}(t)\,, \qquad t > 0\,, \tag{2.7}$$

which follows immediately from (2.5). We define the Hardy-Littlewood relation $f_1 \prec f_1$ between two functions on A to mean

$$\int_0^t f_1^*\, du \leq \int_0^t f_2^*\, du\,, \qquad t > 0\,. \tag{2.8}$$

This is only a quasi-order, for the relation $f_1 \prec f_2$ is not additive. If $|f_1| \leq |f_2|$, then of course $f_1 \prec f_2$, but the converse is not true. If $f_1 \prec f_2$, $f_1, f_2 \geq 0$, if $A = [0, \infty)$ and if g is decreasing and non-negative, then

$$\int_0^\infty f_1^* g\, dt \leq \int_0^\infty f_2^* g\, dt\,. \tag{2.9}$$

Inequality (2.7) may be rewritten as: $(f+g)^* \prec f^* + g^*$.

A Banach function space of measurable functions on A which satisfies $X \subset L_1(A) + L_\infty(A)$ and contains characteristic functions of subsets of A of finite measure is a *rearrangement-invariant space* if $f \in X$ and $|g| \sim |f|$ for a measurable g implies that $g \in X$ and that $\|g\|_X = \|f\|_X$. Then even more is true:

Theorem 2.2. *In a rearrangement-invariant space* X, $f \in X$ *and* $g \prec f$ *imply that also* $g \in X$ *and that* $\|g\|_X \leq \|f\|_X$.

Of course, not all Banach function spaces are rearrangement invariant. The space $L_p(w)$ with the norm $\|f\| = (\int_A w|f|^p\, dx)^{1/p}$, $1 \leq p < \infty$, in general, is not.

In the next section, we shall give some examples of rearrangement-invariant spaces.

There is a standard way to construct rearrangement-invariant Banach function spaces X on A. Let $G = \{g\}$ be a family of decreasing non-negative functions on $\mathbb{R}_+$. We say that $f \in X$ if

$$\|f\|_X := \sup_{g \in G} \int_{\mathbb{R}_+} f^* g\, dx < \infty\,. \tag{2.10}$$

This is called the Luxemburg representation of the norm, which defines a rearrangement-invariant Banach function space (with the Fatou property). Actually, *all rearrangment-invariant Banach function spaces* X *can be obtained in this way.*

We offer some examples of spaces.

Let ϕ be a decreasing non-negative locally integrable function on $A := \mathbb{R}_+$, let $1 \leq q < \infty$. We define the Lorentz spaces $\Lambda := \Lambda(\phi, q)$ and $M := M(\phi, q)$ (see Lorentz [1950] and the book, Lorentz [A-1953]) of functions f on A with the finite norms

$$\|f\|_{\Lambda(\phi,q)} = \left\{ \int_0^\infty \phi(t) f^*(t)^q\, dt \right\}^{1/q} \tag{2.11}$$

$$\|f\|_{M(\phi,q)} = \sup_{c>0} \left\{ \frac{1}{\Phi(c)} \int_0^c f^*(t)^q\, dt \right\}^{1/q}, \tag{2.12}$$

where $\Phi(c) = \int_0^c \phi(t)\, dt$. Using the fact that $(f+g)^* \prec f^* + g^*$, formula (2.9) and Minkowski's inequality, we obtain that the expressions (2.11), (2.12) are subadditive, so that Λ and M are rearrangement-invariant Banach function spaces. The space $\Lambda(\phi, q)$ is separable, the space $M(\phi, q)$ is not – it resembles somewhat the space L_∞. Of particular importance is the case $\phi(t) = t^\alpha$, $\alpha = q/p - 1$. In this way we obtain a two parameter family of Lorentz spaces $L_{p,q}$ consisting of all functions f for which the expression

$$\|f\|^*_{p,q} = \left(\int_0^\infty \left[t^{1/p} f^*(t)\right]^q \frac{dt}{t}\right)^{1/q} \tag{2.13}$$

is finite. It follows from the discussion above that for $1 \le q \le p < \infty$, $L_{p,q}$ is a rearrangement-invariant Banach function space with the norm (2.13); if $q > p$, (2.13) is not a norm. This is the *original definition* of the $L_{p,q}$ spaces. We shall use instead the definition of Calderón (3.7) of the next section. It has the great advantage to work also if $q > p$; if $q \le p$, the norms (2.13) and (3.7) are equivalent if p does not approach 1.

We must briefly mention the Orlicz spaces L^Ψ, which correspond to a function $\Psi(u) = \int_0^u \psi(t)\,dt$, where $\psi(t) \ge 0$ is increasing, with $\psi(0) = 0$. (For $\psi(u) = pu^{p-1}$, $p \ge 1$, one gets the L_p spaces.) One cannot define L^Ψ by the requirement that for f on $[0, b)$,

$$M^\Psi(f) := \int_0^b \Psi(|f(t)|)\,dt < \infty\,,$$

because this set of f would not be, in general, linear. The correct requirement is that $M^\Psi(af) \le 1$ for some $a > 0$. See Bennett and Sharpley [B-1988, p. 265] for details.

§ 3. Hardy's Inequalities and the (θ, q)-quasi-norms

Another approach to concrete rearrangement-invariant spaces is based on Hardy's inequalities. They are applications of integration with respect to Haar measure dx/x on $\mathbb{R}_+$.

Theorem 3.1 (Hardy). *Let $\theta > 0$, $1 \le q \le \infty$, then for each positive measurable function ϕ on $\mathbb{R}_+$,*

$$\int_0^\infty \left[t^{-\theta} \int_0^t \phi(s)\frac{ds}{s}\right]^q \frac{dt}{t} \le \frac{1}{\theta^q} \int_0^\infty \left[t^{-\theta}\phi(t)\right]^q \frac{dt}{t}\,, \tag{3.1}$$

$$\int_0^\infty \left[t^{\theta} \int_t^\infty \phi(s)\frac{ds}{s}\right]^q \frac{dt}{t} \le \frac{1}{\theta^q} \int_0^\infty \left[t^{\theta}\phi(t)\right]^q \frac{dt}{t}\,. \tag{3.2}$$

For $q = \infty$, the integral is replaced by the L_∞ norm. For example, (3.1) *becomes*

$$\sup_{t>0}\left\{t^{-\theta}\int_0^t \phi(s)\frac{ds}{s}\right\} \le \frac{1}{\theta}\operatorname{ess\,sup}_{t>0}\left(t^{-\theta}\phi(t)\right). \tag{3.3}$$

Proof. The second inequality follows from the first upon replacing t by $1/t$ and s by $1/s$. For $q = 1$, the right and the left sides of (3.1) are identical, as a change of integration shows. For $q = \infty$, let M be the right hand supremum in (3.3). Then,

$$\int_0^t \phi(s)\frac{ds}{s} \le \int_0^t M s^{\theta-1}\,ds = \frac{M}{\theta}t^\theta$$

and (3.3) follows.

It remains to prove (3.1) for $1 < q < \infty$. We apply Hölder's inequality with exponents $p, q, 1/q + 1/p = 1$ to

$$\int_0^t \phi(s)\frac{ds}{s} = \int_0^t s^{-\lambda}\phi(s)s^{\lambda-1}\,ds\,.$$

The elementary integrals that appear below will converge for all λ with $1/q < \lambda < \theta + 1/q$, but the best constant in (3.1) is achieved by $\lambda = 1/q + \theta/p$. For this value of λ, we have $(\lambda - 1)p = \theta - 1$, and we obtain

$$\int_0^t \phi(s)\frac{ds}{s} \le \left(\int_0^t s^{-\lambda q}\phi(s)^q\,ds\right)^{1/q} \theta^{-1/p}t^{\theta/p}\,.$$

Changing the order of integration and using the relation $-\theta q + \theta q/p = -\theta$, we obtain for the left side of (3.1) the upper bound

$$\theta^{-q/p}\int_0^\infty t^{-\theta q-1}t^{\theta q/p}\int_0^t s^{-\lambda q}\phi(s)^q\,dsdt = \theta^{-q}\int_0^\infty s^{-\lambda q-\theta}\phi(s)^q\,ds\,.$$

Since $\lambda q + \theta = \theta q + 1$, this is the right side of (3.1). □

An interesting special case of (3.1) is when $\theta = (q-1)/q$, $v(t) = \phi(t)/t$, when we obtain the original inequality of Hardy (Hardy, Littlewood, Pólya [B-1964, p. 240]) for functions $v(t) \ge 0$, $1 < q \le \infty$:

$$\int_0^\infty \left(\frac{1}{t}\int_0^t v(s)\,ds\right)^q dt \le \left(\frac{q}{q-1}\right)^q \int_0^\infty v(t)^q\,dt\,.$$

A similar statement holds for inequality (3.2).

Another method of constructing new spaces, due to Calderón [1966] and Brudnyi and Krugljak [1981] uses, in constrast to (2.10), the functions f^{**} rather than f^*. We start with a functional $\Phi(w)$, $0 \le \Phi(w) \le \infty$, defined for at least each nonnegative concave function w on $\mathbb{R}_+$. We call Φ *admissible* if it has the properties

$$\begin{cases} \text{(i)} & \Phi(w) = 0 \text{ if and only if } w = 0\,,\\ \text{(ii)} & \Phi(aw) = a\Phi(w)\,, \text{ if } a \ge 0\,,\\ \text{(iii)} & w_1 \le w_2 \text{ implies } \Phi(w_1) \le \Phi(w_2)\,,\\ \text{(iv)} & \Phi(\Sigma_i w_i) \le \Sigma_i \Phi(w_i)\,,\\ \text{(v)} & \Phi(w_0) < \infty \text{ for the function } w_0(t) := \min(1, t)\,, \end{cases} \tag{3.4}$$

where the subadditivity condition (iv) is for all countable sums.

Admissible functionals can be used to define functions spaces. We let $L(\Phi)$ be the set of all functions $f \in L_1(A) + L_\infty(A)$ for which

$$\|f\|_{L(\Phi)} := \Phi\left(\int_0^t f^*(s)\,ds\right) \tag{3.5}$$

is finite.

Theorem 3.2. *For an admissible functional Φ, the space $L(\Phi)$, with the norm (3.5), has all properties of a rearrangment-invariant space (except that Fatou's property does not necessarily hold).*

Proof. Properties (i)–(iv) insure that (3.5) is a norm, and (iv) can be used to show that $L(\phi)$ is complete. Property (v) guarantees that each characteristic function of a set of finite measure belongs to $L(\Phi)$. □

As an example of a functional Φ which generates useful norms and quasi-norms we take with $\theta > 0$, $0 < q \le \infty$,

$$\|w\|_{\theta,q} := \Phi_{\theta,q}(w) := \begin{cases} \left(\int_0^\infty [t^{-\theta} w(t)]^q \frac{dt}{t}\right)^{1/q}, & 0 < q < \infty \\ \operatorname{ess\,sup}(t^{-\theta} w(t)), & q = \infty. \end{cases} \tag{3.6}$$

We assume here that $w(t)$ is a non-negative function defined on $\mathbb{R}_+$, which need not be concave.

Now if $0 < \theta < 1$, $1 \le q \le \infty$, then conditions (i)–(v) are satisfied. Indeed, properties (i)–(iii) are trivial and (iv) follows from the subadditivity of $\|\cdot\|_q$ applied to the functions $t^{-\theta-1/q} w_i$. Property (v) follows by a simple computation.

If we take $\theta = 1 - 1/p$, $w(t) = \int_0^t f^*(s)\,ds = tf^{**}(t)$, the norm (3.6) becomes

$$\|f\|_{L_{p,q}} := \|f\|_{p,q} := \begin{cases} \left(\int_0^\infty [t^{1/p} f^{**}(t)]^q \frac{dt}{t}\right)^{1/q}, & 1 \le q < \infty \\ \operatorname{ess\,sup}_{t>0}(t^{1/p} f^{**}(t)), & q = \infty. \end{cases} \tag{3.7}$$

This is *Calderón's definition of the spaces $L_{p,q}$, which we shall use from now on.* Unlike (2.13), the new norms are defined also for $p < q$.

Proposition 3.3. *For $1 < p < \infty$ and $1 \le q < \infty$, the norms (2.13) and (3.7) are equivalent*

$$\|f\|^*_{p,q} \le \|f\|_{p,q} \le p'\|f\|^*_{p,q}, \qquad p' = \frac{p}{p-1}. \tag{3.8}$$

For $1 < p \le \infty$, $1 \le q \le \infty$, $L_{p,q}$ is a rearrangement-invariant Banach function space with norm (3.7).

Proof. The first inequality (3.8) follows from $f^* \le f^{**}$, the second follows from the Hardy inequality (3.1) by taking $\theta = 1 - 1/p$, $\phi(t) = tf^*(t)$. It is easy to derive Fatou's property for the norm (3.7). Therefore the last statement follows from Theorem 3.2. □

We shall give further properties and applications of the quasi-norms (3.6), assuming only that $w(t) \ge 0$ is monotone on $\mathbb{R}_+$. The first remark is that the integral in (3.6) is $= \sum_{k\in\mathbb{Z}} \int_{2^{-k-1}}^{2^{-k}} t^{-\theta q} w(t)^q \frac{dt}{t}$. Let

$$a_k = w(2^{-k}), \quad k \in \mathbb{Z}, \quad \mathbf{a} = (a_k). \tag{3.9}$$

Now the value of $w(t)$ in the last integral lies between a_k and a_{k+1}, while $\int_{2^{-k-1}}^{2^{-k}} t^{-\theta q} \frac{dt}{t}$ is between $2^{\theta q k} \log 2$ and $2^{\theta q(k+1)} \log 2$. Therefore the quasi-norm (3.6) is equivalent to

$$\|\mathbf{a}\|_{\theta,q} := \begin{cases} \left(\sum_{k\in\mathbb{Z}} (2^{k\theta} a_k)^q\right)^{1/q}, & 1 \le q < \infty \\ \sup_{k\in\mathbb{Z}}(2^{k\theta} a_k), & q = \infty. \end{cases} \tag{3.10}$$

The discrete quasi-norm (3.10) has its own interest and uses. We shall prove with its help that the *Hardy's inequalities of Theorem 3.1 are valid even for $0 < q \le \infty$, provided the function ϕ is monotone.*

Lemma 3.4 (The Discrete Hardy Inequality). *Let $\mathbf{a} = (a_k)$, $\mathbf{b} = (b_k)$, $k \in \mathbb{Z}$ be two positive sequences and let for some $C_0 > 0$, $\mu > 0$,*

$$b_k \le C_0 \left(\sum_{j=k}^\infty a_j^\mu\right)^{1/\mu}, \qquad k \in \mathbb{Z}. \tag{3.11}$$

Then for all $\theta > 0$, $q > 0$,

$$\|\mathbf{b}\|_{\theta,q} \le C_0 C \|\mathbf{a}\|_{\theta,q}, \qquad C = C(\theta, q). \tag{3.12}$$

Proof. If (3.11) holds for some μ then it holds for all smaller values of μ because the l_μ norms get larger as μ gets smaller. Therefore, we can assume $\mu < q$. We take $0 < \beta < \theta$ and write $a_j = 2^{j\beta} a_j 2^{-j\beta}$ in (3.11). Then Hölder's inequality with exponents q/μ, r/μ, which satisfy $\mu/q + \mu/r = 1$, gives for each $k \in \mathbb{Z}$,

$$\begin{aligned} b_k &\le C_0 \left(\sum_{j=k}^\infty [2^{j\beta} a_j]^q\right)^{1/q} \left(\sum_{j=k}^\infty 2^{-j\beta r}\right)^{1/r} \\ &\le C C_0 2^{-k\beta} \left(\sum_{j=k}^\infty [2^{j\beta} a_j]^q\right)^{1/q}. \end{aligned}$$

Hence, the left side of (3.12) does not exceed

$$CC_0\Big(\sum_{k\in\mathbb{Z}}\sum_{j=k}^{\infty} 2^{k(\theta-\beta)q}2^{j\beta q}a_j^q\Big)^{1/q}$$

and an interchange of summation shows that this is not larger than the right side of (3.12). □

Remark. One can (with a similar proof) replace (3.11) in Lemma 3.3 by the assumption that

$$b_k \le C_0 2^{-k\lambda}\Big\{\sum_{j=-\infty}^{k}(2^{j\lambda}a_j)^\mu\Big\}^{1/\mu}, \qquad k\in\mathbb{Z}. \tag{3.13}$$

Then (3.12) holds for all $0<\theta<\lambda$.

Theorem 3.5. *For a non-negative monotone function ϕ, the inequalities (3.1), (3.2) of Theorem 3.1 remain valid for $\theta>0$, $0<q<1$, provided the factor $1/\theta^q$ is replaced by some constant $C:=C(\theta,q)$.*

Proof. Because of the change of variables mentioned in the proof of Theorem 3.1, it is sufficient to prove (3.1); we shall assume that ϕ is increasing; the other case is quite similar. If $a_k:=\phi(2^{-k})$, $k\in\mathbb{Z}$, then

$$b_k := \int_0^{2^{-k}}\phi(s)\frac{ds}{s} = \sum_{j=k}^{\infty}\int_{2^{-j-1}}^{2^{-j}}\phi(s)\frac{ds}{s} \le (\log 2)\sum_{j=k}^{\infty}a_j\,.$$

We can apply Lemma 3.4 and obtain (3.12). The outer integral on the left in (3.1) we write as a sum of integrals over the intervals $[2^{-k-1},2^{-k}]$. In this way we find that it does not exceed

$$(\log 2)2^{\theta q}\sum_{k\in\mathbb{Z}}\Big[2^{k\theta}\int_0^{2^{-k}}\phi(s)\frac{ds}{s}\Big]^q = (\log 2)2^{\theta q}\|\mathbf{b}\|_{\theta,q}^q\,.$$

Similarly, the right integral in (3.1) is

$$\ge (\log 2)2^{-\theta q}\sum_{k\in\mathbb{Z}}[2^{k\theta}a_k]^q = (\log 2)2^{-\theta q}\|\mathbf{a}\|_{\theta,q}^q\,.$$ □

§ 4. Linear Operators. Interpolation of Operators

Operators of different types are a tool and a subject of approximation theory. An operator $g=U(f)$ is a map of a space X into another space Y. It is linear, if for all scalars a_1,a_2, $U(a_1f_1+a_2f_2)=a_1U(f_1)+a_2U(f_2)$, $f_1,f_2\in X$.

If Y is a subspace of a linear normed space X, and $f\in X$, then we call

$$E(f):=E(f,Y):=\operatorname{dist}(f,Y)=\inf_{g\in Y}\|f-g\| \tag{4.1}$$

the *error of approximation* of f by Y. A $g\in Y$ for which the infimum is attained, is *an element of best approximation to f from Y*. If for each $f\in X$ there is a unique element of best approximation, we have *the operator of best approximation* $P(f)$. Unfortunately these operators do not have good properties (they are usually not linear, see §1 of Chapter 3). What is more important, there only very seldom exist usable representations of P. Therefore one usually replaces P by simpler operators – for instance by linear integral operators (4.7) which may provide good, but not necessarily best approximation.

A continuous linear operator U, mapping a Banach space X into itself, is of finite rank n if its range $Y_n=\{Uf : f\in X\}$ is a linear space of dimension n. If $g_1,\dots,g_n$ are a basis of Y_n, we can write

$$Uf=\sum_{i=1}^{n}c_i(f)g_i\,. \tag{4.2}$$

Applying to both sides of (4.2) a linear functional which is orthogonal to all but one g_i, one proves that the coefficients $c_i(f)$ are continuous linear functionals for $f\in X$. In many cases for a separable space X, the identity operator $If=f$ can be approximated by operators U_n of finite rank:

$$\|f-U_nf\|\to 0 \quad\text{for all}\quad f\in X\,. \tag{4.3}$$

The *norm* of a linear operator U which maps a Banach space X into another such space Y is given by

$$\|U\|:=\|U\|_{X\to Y}:=\sup_{\|f\|_X\le 1}\|U(f)\|_Y\,; \tag{4.4}$$

the operator U is continuous if and only if it is *bounded*, that is, it satisfies $\|U\|<\infty$. The importance of the estimation of the norms from above is clear from the *Banach-Steinhaus theorem*, which gives necessary and sufficient conditions for the convergence

$$U_n(f)\to U(f), \qquad f\in X \tag{4.5}$$

of a sequence of linear bounded operators U_n. The conditions are:

(a) convergence $U_n(g)\to U(g)$ for a set of $g\in X$ whose linear combinations are dense in X;
(b) boundedness of the norms, $\|U_n\|\le M$, $n=1,2,\dots$.

In some cases, for positive operators U_n, (see Theorem 3.1 in Chapter 1), one has to test (a) only for a finite set of g.

An operator U of X into itself is a *projection* onto a linear subspace Y, if $Uf\in Y$ for all $f\in X$, and $Uf=f$ for $f\in Y$. If U_n are projections of X onto finite dimensional subspaces X_n, $n=1,2,\dots$, $X_1\subset X_2\subset\cdots$, and if the closure of $\cup X_n$ spans X, then in the Banach-Steinhaus theorem one has to check only (b). The following simple remark about projections is often useful:

Proposition 4.1 (Lebesgue's Lemma). *If U is a projection of X onto Y, and if $E(f)$ is the error of approximation of f by Y, then*

$$\|f - Uf\| \le (1 + \|U\|)E(f). \tag{4.6}$$

Indeed, if $g \in Y$, then $Ug = g$ and therefore

$$\|f - Uf\| \le \|f - g\| + \|U(f-g)\| \le (1 + \|U\|)\|f - g\|.$$

and (4.6) follows. □

Thus, if the norm $\|U\|$ is not large, there is no serious loss of precision if we replace the best approximant g by Uf. For example, the n-th partial sum of the Fourier series is an operator $s_n(f)$ with norm $\le C \log n$ in $C(\mathbb{T})$. (See Theorem 2.1 of Chapter 9.) Hence, if the Fourier series of f converges well, there can be only very little advantage in replacing $s_n(f)$ by the trigonometric polynomial T_n of best uniform approximation to f.

Of particular importance are integral operators

$$U(f) := U(f,x) := \int_A K(x,t)f(t)\,dt, \qquad f \in X \tag{4.7}$$

where A is a finite or infinite interval of $\mathbb{R}$ or $\mathbb{T}$, and sums (which are relevant if $f \in C(A)$):

$$U(f,x) = \sum_{k=1}^{n} a_k(x) f(x_k). \tag{4.8}$$

We assume that the kernel $K(x,y)$ in (4.7) is continuous on A^2 and define

$$\begin{cases} M_\infty := \operatorname*{ess\,sup}_{x \in A} \int_A |K(x,t)|\,dt \\ M_1 := \operatorname*{ess\,sup}_{t \in A} \int_A |K(x,t)|\,dx. \end{cases} \tag{4.9}$$

For compact A, the ess sup is to be replaced by maximum. In this case, for continuous f, the function $U(f)$ is also continuous on A.

The following theorem describes the norm of the operators (4.7) on L_1 and on L_∞:

Theorem 4.2. (i) *The condition $M_\infty < \infty$ is necessary and sufficient for the operator* (4.7) *to map L_∞ into itself, and M_∞ is its norm. It is also the norm of U on the space $L_\infty \cap C(A)$ of bounded continuous functions.* (ii) *The condition $M_1 < \infty$ is necessary and sufficient for U to map L_1 into itself; the norm of U is then equal to M_1.*

Proof. The *sufficiency* of the conditions and the inequalities $\|U\|_\infty \le M_\infty$, $\|U\|_1 \le M_1$ are easily proved. If $g = U(f)$, then

$$|g(x)| \le \int_A |K(x,t)|\,dt\,\|f\|_\infty \le M_\infty \|f\|_\infty, \text{ a.e.}$$

hence the norm $\|U\|_\infty \le M_\infty$ is finite. Similarly for the $L_1 \to L_1$ map; the estimate $\|U\|_1 \le M_1$ follows from Fubini's theorem and Minkowski's inequality.

In the inverse direction, we first assume that A is compact; let for example $A = [a,b]$.

(i) If U maps L_∞ into itself, we select $x_0 \in A$ so that $M_\infty = \int_A |K(x_0,t)|\,dt$. With $h(t) :=$ sign $K(x_0,t)$, we have then $U(h,x_0) = M_\infty$. Since $U(h,x)$ is continuous and $\|h\| \le 1$, we deduce $\|U\| \ge M_\infty$. The same result holds for the integral as a map of $C(A)$ into itself. Since $K(x_0,\cdot)$ is continuous, for a given $\varepsilon > 0$, there is a continuous function f which agrees with h on the set $E := \{t : |K(x_0,t)| \ge \varepsilon\}$, and satisfies $|f(t)| \le 1$, $t \notin E$. Then,

$$U(f,x_0) = \int_A h(t)K(x_0,t)\,dt + \int_{A\setminus E} (f(t) - h(t))K(x_0,t)\,dt \ge M_\infty - 2\varepsilon|A|.$$

Letting $\varepsilon \to 0$, we obtain (i).

(ii) If (4.7) maps L_1 into itself, we select $t_0 \in [a,b]$ for which

$$M_1 = \int_A |K(x,t_0)|\,dx.$$

Let $\varepsilon > 0$ be given. We select $0 < \delta < b - a$ so that $|t - s| < \delta$ implies $|K(x,t) - K(x,s)| < \varepsilon$ and take an interval $I \subset A$ of length δ containing t_0. For the function $f := \delta^{-1}\chi_I$ and $g := U(f)$, we have

$$\begin{aligned} \|g\|_1 = \frac{1}{\delta}\int_A \left|\int_I K(x,t)\,dt\right| dx &\ge \frac{1}{\delta}\int_A \left|\int_I K(x,t_0)\,dt\right| dx \\ &\quad - \frac{1}{\delta}\int_A \int_I |K(x,t) - K(x,t_0)|\,dt\,dx. \end{aligned}$$

The first integral is $= M_1$, while the second integral does not exceed $\varepsilon(b-a)$. Hence $\|g\|_1 \ge M_1 - \varepsilon(b-a)$, and since $\|f\|_1 = 1$, we get $\|U\| = M_1$.

For unbounded A, the proof is completed by a limiting process. We work this out for $A = \mathbb{R}$ in case (i). We first note that

$$M_\infty^{(n)} := \max_{-n \le x \le n} \int_{-n}^{n} |K(x,t)|\,dt \to M_\infty \quad \text{as } n \to \infty.$$

Indeed, the continuous functions $\phi_n(x) = \int_{-n}^{n} |K(x,t)|\,dt$ form an increasing sequence and hence converge pointwise to $\phi(x) = \int_A |K(x,t)|\,dt$, hence ess sup $\phi_n \to$ ess sup $\phi = M_\infty$. Let f_n be the continuous function constructed for the compact set $[-n,n]$. Without loss of generality, we can assume that $f_n(-n) = f_n(n) = 0$. We extend f_n by zero outside of $[-n,n]$ and get $\|Uf_n\|_\infty \ge M_\infty^{(n)} - \delta$, for arbitrarily small $\delta > 0$, hence $\|U\| \ge M_\infty$. □

On $\mathbb{R}$ and $\mathbb{T}$ many integral operators are given by a convolution

$$(Uf)(x) := (K * f)(x) := \int_A K(x-t)f(t)\,dt \tag{4.10}$$

with continuous K. In this case, the norm of U on L_∞ and on L_1 is given simply by $\|U\| = \int_A |K|\,dt$.

In spaces other than L_1, L_∞ it is much more difficult to find exactly the norm of the integral operator (4.7). We mention only that as a map of L_2 into itself, we have

$$\|U\|_2^2 \le \int_A \int_A |K(x,t)|^2\,dx\,dt\,. \tag{4.11}$$

(If the integral on the right is finite, (4.7) is called a Hilbert-Schmidt operator.) This is obtained by applying the Cauchy inequality to the integral (4.7) and then taking the L_2-norm of the result. The exact expression for $\|U\|_2$ is more complex. For a Hilbert-Schmidt operator U with a symmetric kernel K, $\|U\|_2 = |\lambda|$, where λ is the largest in modulus eigenvalue of U.

Another source of estimates of norms of operators are *interpolation theorems*. An interpolation theorem is a statement about three pairs of spaces of the following kind. If a linear operator maps a space X_i into Y_i with norm M_i, $i = 1, 2$, then U (or a natural extension of it) maps also X into Y. The norm of the latter map can be often estimated in terms of M_1, M_2. If this happens, then (X, Y) is an *interpolation pair between* (X_i, Y_i), $i = 1, 2$. In particular, if $X_i = Y_i$, $i = 1, 2$, $X = Y$, then X is an *interpolation space between* X_1 *and* X_2.

The famous *interpolation theorem of Riesz-Thorin* deals with L_p spaces. We write $p = 1/\alpha$, then the range $1 \le p \le \infty$ becomes $0 \le \alpha \le 1$, and a pair L_p, L_q is represented by a point (α, β) in the square $Q : 0 \le \alpha, \beta \le 1$. The theorem asserts that we can interpolate along any straight segment contained in Q. In this theorem, the L_p are *complex spaces*.

Theorem 4.3. *Let $1 \le p_1, p_2, q_1, q_2 \le \infty$, $0 \le \theta \le 1$ and let p, q be defined by*

$$\frac{1}{p} = (1-\theta)\frac{1}{p_1} + \theta\frac{1}{p_2}\,, \qquad \frac{1}{q} = (1-\theta)\frac{1}{q_1} + \theta\frac{1}{q_2}\,. \tag{4.12}$$

If U is a linear bounded operator that maps $L_{p_i} \to L_{q_i}$ with norm M_i, $i = 1, 2$, then U can be extended to a linear bounded operator $L_p \to L_q$ with norm

$$\|U\|_{L_p \to L_q} \le M_1^{1-\theta} M_2^{\theta}\,. \tag{4.13}$$

(In other words, $\log \|U\|_{L_{p,q}}$ is a convex function of θ on the interval $0 \le \theta \le 1$.)

This theorem has been first proved by M. Riesz for the "lower" triangle $0 \le \beta \le \alpha \le 1$ by real methods but the full statement requires tools from complex variable theory. (See Bennett and Sharpley [B-1988, p. 196].)

The relation (4.13) is exact: equality can happen for each θ, $0 < \theta < 1$. A special case is when $p_1 = q_1 = 1$, $p_2 = q_2 = \infty$. Then $p = q$, and the theorem implies: each operator mapping $L_1 \to L_1$ and $L_\infty \to L_\infty$ also maps $L_p \to L_p$, $1 \le p \le \infty$, with norm $\le \max(M_1, M_2)$.

Instead of Theorem 4.3, we shall use a theorem of Hardy, Littlewood and Pólya, which, restricted to the L_p-spaces, yields only the special statement above (and not (4.13)). Its advantages are a simple real proof, and the possibility of replacing the L_p by arbitrary rearrangement-invariant spaces.

If U is a linear operator that maps $L_1(A) \to L_1(A)$ and $L_\infty(A) \to L_\infty(A)$, it can be uniquely extended onto $L_1(A) + L_\infty(A)$ by means of the formula

$$Uf = Uf_1 + Uf_2 \quad \text{for} \quad f = f_1 + f_2\,, \ f_1 \in L_1\,, \ f_2 \in L_\infty\,. \tag{4.14}$$

This follows at once from the fact that representation (4.14) and $f = f_1' + f_2'$, $f_1' \in L_1$, $f_2' \in L_\infty$ imply that $g = f_1 - f_1' = -(f_2 - f_2') \in L_1 \cap L_\infty$, so that the two representations produce the same value for Uf.

Theorem 4.4 (Hardy, Littlewood, Pólya). *Let U be a linear bounded operator from $L_1(A)$ into itself and from $L_\infty(A)$ into itself with norms not exceeding M. Then its natural linear extension onto $L_1(A) + L_\infty(A)$ (also denoted by U) has the property*

$$Uf \prec Mf\,, \qquad f \in L_1 + L_\infty\,; \tag{4.15}$$

it maps each rearrangement invariant space $X \subset L_1 + L_\infty$ into itself with norm $\le M$.

For matrix operators on finite dimensional sequence spaces this theorem appears in the book [B-1968] of Hardy, Littlewood, Pólya, for integral operators in Lorentz [A-1953], for the general case (stated in a different way) in O'Neil-Weiss [1963].

Proof. For given $f \in L_1 + L_\infty$ and each finite a, $0 < a \le |A|$, by (2.5) we can find measurable sets $E_1 \subset A$, $E_2 \subset A$ for which $|E_1| = |E_2| = a$ and $\int_0^a f^*\,dt = \int_{E_1} |f|\,dx$, $\int_0^a (Uf)^*\,dt = \int_{E_2} |Uf|\,dx$. Let $c = f^*(a)$. We define the c-truncation of f on A by putting $f_c(x) := f(x)$ if $|f(x)| \le c$, $f_c(x) := c \operatorname{sign} f(x)$ otherwise. Then, using both $\|U\|_1 \le M$ and $\|U\|_\infty \le M$, we get

$$\begin{aligned} \int_0^a (Uf)^*\,dt &= \int_{E_2} |Uf|\,dx \\ &\le \int_{E_2} |U(f - f_c)(x)|\,dx + \int_{E_2} |Uf_c|\,dx \\ &\le M \int_A |f - f_c(x)|\,dx + Mc|E_2|\,. \end{aligned}$$

Outside of E_1, $f(x) = f_c(x)$, and on E_1, $f(x)$ and $f_c(x)$ are of the same sign. Therefore the last expression is

$$= M\int_{E_1}|f|\,dx - M\int_{E_1}|f_c|\,dx + Mc|E_1|$$
$$= M\int_{E_1}|f|\,dx = M\int_0^a f^*(t)\,dt\,.$$

This proves (4.15). The remaining statement follows from this and Theorem 2.2. □

The following result will be very often used:

Theorem 4.5. *Let X be a rearrangement-invariant function space on A, let $K(x,t)$ be measurable on A^2. If*

$$\int_A |K|\,dx \le M \text{ a.e.}, \quad \int_A |K|\,dt \le M \text{ a.e.}, \tag{4.16}$$

then the integral

$$U(f,x) = \int_A K(x,t)f(t)\,dt\,, \qquad f\in X \tag{4.17}$$

is defined a.e. on A and represents a continuous linear operator on X of norm $\le M$.

Proof. Because of (4.16), the integral exists for almost all x and for all $f\in L_\infty(A)$ or $f\in L_1(A)$ (for the latter statement we use Fubini's theorem), hence for all $f\in X\subset L_\infty(A)+L_1(A)$. The integral coincides with the operator of Theorem 4.4 obtained by extension since it does for any characteristic function of a set of finite measure. Hence (4.17) is a bounded linear operator of norm $\le M$. □

Corollary 4.6. *The norm of the operator* (4.10) *on any rearrangement invariant function space X on A does not exceed $\int_A |K|\,dt$.*

Example. Let f_h, $0<h<b-a$ be the Steklov function of $f\in L_1[a,b]$,

$$f_h(x) := \begin{cases} h^{-1}\int_x^{x+h} f(t)\,dt\,, & a\le x\le b-h\\ = f_h(b-h) & b-h\le x\le b\,.\end{cases} \tag{4.18}$$

This is an integral operator of the type (4.17) with a kernel $K(x,t)$ that satisfies (4.16) with $M=2$. For continuous f on $[a,b]$, $f_h\to f$ uniformly. By the Banach-Steinhaus theorem, for each rearrangment invariant function space X on $[a,b]$, for which continuous functions are dense, we have

$$\|f-f_h\|_X\to 0\,,\quad h\to 0+\,,\qquad f\in X\,. \tag{4.19}$$

§ 5. Spaces of Differentiable Functions: Sobolev Spaces

Some of the basic spaces of smooth functions are related to differentiation. As usual we restrict the domain of definition to be $A=\mathbb{R}:=(-\infty,\infty)$, or $\mathbb{R}_+:=[0,\infty)$, the circle $\mathbb{T}$, or an interval $[a,b]$. For a Banach space X of functions defined on A, we denote by $W^r(X)$, $r=1,2,\dots$, the collection of functions for which $f^{(r-1)}$ is absolutely continuous and $f^{(r)}\in X$. The seminorm and norm of $W^r(X)$ are

$$|f|_{W^r(X)} := \|f^{(r)}\|_X\,,\quad \|f\|_{W^r(X)} := \|f\|_X + |f|_{W^r(X)}\,. \tag{5.1}$$

The choice $X=L_p(A)$, $1\le p\le\infty$, yields the *Sobolev* spaces $W_p^r := W^r(L_p)$. If $X=C(A)$, we get the *spaces* $C^r(A):=W^r(C(A))$ *of r-times continously differentiable functions on A*; the space $C^\infty(A)$ consists of *infinitely differentiable functions* on A. By $C_0^r(A)$, $r=1,\dots,\infty$, we denote the subspace of $C^r(A)$ consisting of functions with compact support, which is interior to A.

Results for non-compact intervals follow often from their finite interval analogues by using partitions of unity. A finite or infinite sequence of functions (ϕ_j) is a *partition of unity* for A if $\sum\phi_j(x)=1$, $x\in A$.

To construct useful partitions of unity, we shall use the following simple lemma.

Lemma 5.1. *For each finite interval $I=[a,b]$ and for $0<\delta<(b-a)/2$ there exists a function $\phi:=\phi_\delta\in C_0^\infty(\mathbb{R})$ which is increasing on $[a,a+\delta]$, decreasing on $[b-\delta,b]$, $=0$ outside of $[a,b]$, $=1$ on $[a+\delta,b-\delta]$, and for which, for some $C_k=C(k,\delta)$, $\|\phi^{(k)}\|_\infty\le C_k|I|^{-k}$, $k=0,1,\dots$. In particular, for each $0<\theta<1$ and each I, there is a $C_k:=C(k,\theta)$ so that for $2\delta\ge\theta|I|$,*

$$\|\phi^{(k)}\|_\infty(\mathbb{R})\le C(k,\theta)|I|^{-k}\,,\qquad k=0,1,\dots$$

Proof. We can assume that $I=[0,1]$; the general case follows from this by a linear substitution. Let $g(x)=e^{-1/x^2}$, $x>0$, and $g(x):=0$, $x\le 0$; plainly $g\in C^\infty(\mathbb{R})$. The function $G(x):=Cg(x)g(\delta-x)$ is non-negative, vanishes outside of $[0,\delta]$ and for properly chosen C has integral one. We can take

$$\phi(x) := \int_{x-c}^x G(u)\,du \quad\text{where}\quad c:=1-\delta\,.$$ □

Theorem 5.2. *Let $\mathbb{R}=\bigcup_{j\in\mathbb{Z}} I_j$ be an infinite union of overlapping intervals $I_j:=[a_j,b_j]$ which for some $0<\theta<1$ satisfy:*

$$a_j<b_{j-1}\le a_{j+1}<b_j\,,\quad b_{j-1}-a_j\ge\frac12\theta(|I_{j-1}|+|I_j|)\,,\qquad j\in\mathbb{Z}\,. \tag{5.2}$$

Then there are functions $\phi_j\in C^\infty$, $j\in\mathbb{Z}$, for which

$$\begin{cases} \text{(i)} & \sum_{j\in\mathbb{Z}}\phi_j(x)\equiv 1\,, \qquad x\in\mathbb{R}\,;\\ \text{(ii)} & 0\le\phi_j\le 1\,,\ \textit{and}\ \phi_j\ \textit{is supported on}\ I_j\,,\quad j\in\mathbb{Z}\,;\\ \text{(iii)} & \|\phi_j^{(k)}\|_\infty\le C(k,\theta)(|I_{j-1}|^{-k}+|I_j|^{-k}+|I_{j+1}|^{-k})\,.\end{cases} \tag{5.3}$$

Proof. For all $j \in \mathbb{Z}$ we define $\delta_j := \frac{1}{2}\theta|I_j|$ and ψ_j to be the function ϕ of Lemma 5.1 for $I = I_j$ and the number θ of (5.2). We have supp $\psi_j = I_j$, $0 \le \psi_j(x) \le 1$. Since $\delta_j + \delta_{j+1} < b_j - a_{j+1}$, the intervals $[a_j + \delta_j, b_j - \delta_j]$ cover $\mathbb{R}$. The function $\Psi = \sum_{j\in\mathbb{Z}} \psi_j$ belongs to C^∞ and satisfies $\Psi(x) \ge 1$, $x \in \mathbb{R}$. Relations (i), (ii) will be satisfied for the functions $\phi_j(x) := \psi_j(x)/\Psi(x)$, $j \in \mathbb{Z}$. For each j and $x \in I_j$, only the terms $\psi_{j-1}(x)$, $\psi_j(x)$, $\psi_{j+1}(x)$ of the sum Ψ are different from zero, hence

$$|\Psi^{(k)}(x)| \le C(k,\theta)(|I_{j-1}|^{-k} + |I_j|^{-k} + |I_{j+1}|^{-k}) =: \Lambda_j(k).$$

Differentiating the relation $\Psi\Psi^{-1} = 1$, we obtain $\Psi^{(k)}\Psi^{-1} + \frac{k}{1}\Psi^{(k-1)}(\Psi^{-1})' + \cdots + (\Psi^{-1})^{(k)} = 0$. By induction, $|(\Psi^{-1})^{(k)}(x)| \le C_k\Lambda_j(k)$, $x \in I_j$, and the Leibniz' formula yields (iii). □

For any subinterval A of $\mathbb{R}$, we obtain from Theorem 5.2 a partition of unity for A by taking those ϕ_j which do not vanish identically on A.

Originally, Sobolev spaces were defined for functions of several variables, using generalized instead of ordinary partial derivatives. For functions of one variable, as we shall show, both ordinary and generalized derivatives produce the same spaces. For a detailed theory of Sobolev spaces see Adams [B-1975].

For two locally integrable functions f, g on A, we say that g is an r-th *generalized derivative* of f on A if

$$(5.4)\qquad \int_A f\phi^{(r)}\,dx = (-1)^r \int_A g\phi\,dx\,, \quad \text{for all } \phi \in C_0^\infty(A).$$

Since ϕ and all its derivatives vanish at the endpoints of A, (5.4) holds in particular if g is an ordinary continuous r-th derivative of f. The following theorem characterizes functions whose r-th generalized derivative is zero.

Theorem 5.3. *Let $A = [a,b]$, $\mathbb{R}$, or $\mathbb{R}_+$ and $r \ge 0$. If $f \in L_{\mathrm{loc}}(A)$ satisfies,*

$$(5.5)\qquad \int_A f\phi^{(r)}\,dx = 0\,, \quad \text{for all } \phi \in C_0^\infty(A),$$

then there is a polynomial P of degree $< r$ such that $f = P$ a.e. on A.

Proof. We assume first that A is a finite interval which we can take to be $[0,1]$ and prove the theorem by induction on r. If (5.5) holds for $r = 0$, we want to show that $f = 0$ a.e. Let $[a,b] \subset A$ and $\delta > 0$. Then the function ϕ_δ of Lemma 5.1 is in $C_0^\infty(A)$ and so $\int_A f\phi_\delta\,dx = 0$. But $\phi_\delta = 1$ on $[a+\delta, b-\delta]$ and $\|\phi_\delta\|_\infty \le 1$ and therefore

$$0 = \lim_{\delta\to 0}\int_A f\phi_\delta\,dx = \int_a^b f(x)\,dx.$$

Thus for any $x \in A$ we have $\int_a^x f(t)\,dt = 0$. Lebesgue's differentiation theorem gives that $f = 0$ a.e.

Suppose now that the theorem has been established for $r = k$ and that f satisfies (5.5) for $r = k+1$ and all ϕ. Let $h, \psi \in C_0^\infty(A)$ and $\int_A h\,dx = 1$. We define $\alpha := \int_A \psi\,dx = (-1)^k\int_A \psi^{(k)}x^k\,dx/k!$ and $\phi := \psi - \alpha h$. Then $\Phi(x) := \int_0^x \phi(u)\,du$ is also in $C_0^\infty(A)$. Hence, with $\beta := (-1)^k\int_A fh^{(k)}\,dx/k!$, we have from (5.4)

$$0 = \int_A f\Phi^{(k+1)}\,dx = \int_A f(\psi^{(k)} - \alpha h^{(k)})\,dx = \int_A \psi^{(k)}(f - \beta x^k)\,dx.$$

Since ψ is an arbitrary function in $C_0^\infty(A)$, by our induction hypothesis $f(x) - \beta x^k = P(x)$ a.e. where P is a polynomial of degree $< k$. We have established the theorem in the case when A is finite.

If A is not finite then for any function f which satisfies (5.5), there is a polynomial P_I of degree $< r$ such that $f = P_I$ a.e. on I for each finite interval I. Covering A by overlapping intervals I, we see that all the P_I are the same and the theorem follows. □

In particular, the generalized derivative of f, if it exists, is unique.

Theorem 5.4. *Let $A = [a,b]$, $\mathbb{R}_+$, or $\mathbb{R}$, and let $r \ge 1$. If $f \in L_{\mathrm{loc}}(A)$ has a generalized r-th derivative $g \in L_{\mathrm{loc}}(A)$, then f can be redefined on a set of measure zero so that $f^{(r-1)}$ is absolutely continuous and $f^{(r)} = g$ a.e. on A.*

Proof. Let G be an r-fold integral of g. Then integration by parts shows that

$$\int_A (f-G)\phi^{(r)}\,dx = \int_A f\phi^{(r)}\,dx - (-1)^r\int_A g\phi\,dx = 0,$$

for each $\phi \in C_0^\infty(A)$. Hence 0 is a generalized r-th derivative of $f - G$ and from Theorem 5.3, $f = G + P$ a.e. where P is a polynomial of degree $< r$. The function $G + P$ has an absolutely continuous $(r-1)$-st derivative and its r-th derivative is g. □

It follows from this that the definition of W_p^r could just as well be given with generalized derivatives. We shall now discuss some inequalities valid for the derivatives of functions $f \in W_p^r(A)$.

A function $f \in W_p^r(A)$ has continuous derivatives of order $k = 0, \dots, r-1$. Therefore, if $c \in A$, we can form the Taylor polynomial $T_{r-1}(x) := T_{r-1}(f,c;x) := f(c) + f'(c)(x-c) + \cdots + f^{(r-1)}(c)(x-c)^{r-1}/(r-1)!$. Integration by parts shows that

$$(5.6)\qquad f(x) - T_{r-1}(x) = \int_c^x f^{(r)}(t)\,\frac{(x-t)^{r-1}}{(r-1)!}\,dt.$$

We shall frequently make use of the following elementary estimates for the remainder $f - T_{r-1}$.

Proposition 5.5. *If $f \in W_p^r(A)$, $A = [a,b]$, and $1 \le p, q \le \infty$, then for the Taylor polynomial $T_{r-1}(x) = T_{r-1}(f,c;x)$, $c \in A$, we have*

$$\|f - T_{r-1}\|_q(A) \le \frac{1}{(r-1)!}|A|^{r-\frac{1}{p}+\frac{1}{q}}\|f^{(r)}\|_p(A)\,. \tag{5.7}$$

Proof. We apply Hölder's inequality in (5.6) and find for $x \in A$

$$|f(x) - T_{r-1}(x)| \le \frac{1}{(r-1)!}|A|^{r-\frac{1}{p}}\|f^{(r)}\|_p(A)\,.$$

We obtain (5.7) by applying the $L_q(A)$ norm to the last inequality. □

If $f \in W_p^r$, then the intermediate derivatives of f exist and we have

Theorem 5.6. *For $r \ge 2$, $1 \le p \le \infty$, and $A = \mathbb{R}, \mathbb{R}_+, [a,b]$, or $\mathbb{T}$, there is a constant C depending only on r such that*

$$u^k\|f^{(k)}\|_p \le C\big(\|f\|_p + u^r\|f^{(r)}\|_p\big)\,, \qquad f \in W_p^r\,,\ k = 0,\dots,r\,, \tag{5.8}$$

where $u > 0$ is arbitrary, except when $A = [a,b]$, in which case $0 \le u \le b-a$. In the last case, we have (5.8) also for $0 \le u \le c$, where $c > 0$ is arbitrary, but with C depending on r and $c/(b-a)$.

Proof. From Taylor's formula (5.6), we have

$$\begin{aligned} f(x+u) &= f(x) + uf'(x) + \cdots + \frac{u^{r-1}}{(r-1)!}f^{(r-1)}(x) \\ &\quad + \int_0^u \frac{(u-t)^{r-1}}{(r-1)!}f^{(r)}(x+t)\,dt \end{aligned} \tag{5.9}$$

which is valid for all $x \in A'$, where $A' := A$ except in the case $A = [a,b]$ when $A' := [a,(a+b)/2]$ and u is arbitrary except in the case $A = [a,b]$ when $0 \le u \le |A|/2$.

The remainder $R_r(x,u)$ in (5.9) is a function of x whose norm in $L_p(A')$ by Minkowski's inequality is equal to $\theta u^r\|f^{(r)}\|_p$, $|\theta| \le 1/r!$. We select fixed numbers $1 =: \lambda_1 < \cdots < \lambda_{r-1} := 2$; then (with the restriction $0 \le u \le |A|/4$ in case $A = [a,b]$),

$$\begin{aligned} &\lambda_s uf'(x) + \cdots + \lambda_s^{r-1}\frac{u^{r-1}}{(r-1)!}f^{(r-1)}(x) \\ &\quad = f(x+\lambda_s u) - f(x) - R_r(x,\lambda_s u)\,, \quad s = 1,\dots,r-1\,. \end{aligned} \tag{5.10}$$

Since the Vandermonde determinant is non-zero, these equations can be solved for $\frac{u^k}{k!}f^{(k)}(x)$. In this way, we obtain

$$u^k\|f^{(k)}\|_p(A') \le C(\|f\|_p + u^r\|f^{(r)}\|_p)\,, \qquad k = 0,\dots,r\,. \tag{5.11}$$

This is (5.8) except when $A = [a,b]$. In this case, a similar inequality holds for $[(a+b)/2, b]$ in place of A' and by addition, we obtain (5.8) for $u \le (b-a)/4$. For other values of u, we derive (5.8) from the case $u = (b-a)/4$ by simply altering the constant C. □

Corollary 5.7. *If $A = \mathbb{T}, \mathbb{R}$, or $\mathbb{R}_+$, there exists constants $C_{r,k}$, $k = 0,\dots,r$, and $r = 1,2,\dots$ with the property*

$$\|f^{(k)}\|_p \le C_{r,k}\|f\|_p^{1-k/r}\|f^{(r)}\|_p^{k/r}\,, \qquad f \in W_p^r(A)\,. \tag{5.12}$$

Proof. One obtains this by taking $u := (\|f\|_p/\|f^{(r)}\|_p)^{1/r}$ in (5.8). If $\|f^{(r)}\|_p = 0$, this is impossible, but then for large u, (5.8) shows that $\|f^{(k)}\|_p = 0$, $0 < k < r$. □

If $r = 2$, instead of several equations in (5.10), we have only one: $uf'(x) = f(x+u) - f(x) - R_2(x,u)$, and therefore

$$u\|f'\|_p \le 2\|f\|_p + u^2\|f''\|_p/2\,. \tag{5.13}$$

Minimizing with respect to u gives

$$\|f'\|_p \le 2\|f\|_p^{1/2}\|f''\|_p^{1/2}\,. \tag{5.14}$$

The important problem of finding the best constants in (5.12) will be discussed in Chapter 5. In particular, we shall see that the best constant in (5.14) for $p = \infty$ is $\sqrt{2}$ instead of 2 (Landau).

The convergence $f_n \to f$ in $W_p^r(A)$ not only implies $\|f_n^{(k)} - f^{(k)}\|_p(A) \to 0$, $0 \le k \le r$ by Theorem 5.6, but also

$$f_n^{(k)} \to f^{(k)}\,,\ 0 \le k < r\,,\ \textit{uniformly on each compact subinterval } B \textit{ of } A. \tag{5.15}$$

Indeed, let $g \in W_p^1$. We have $g(y) = g(x) + \int_x^y g'(u)\,du$, $x, y \in B$, where $|\int_x^y g'\,du| \le \|g'\|_p|B|^{1/p'}$. Applying Minkowski's inequality to the last equation, we get

$$|B|^{1/p}|g(y)| \le \|g\|_p + |B|\,\|g'\|_p\,, \qquad y \in B\,. \tag{5.16}$$

We obtain (5.15) by taking $g := f_n^{(k)} - f^{(k)}$.

In the converse direction, for compact A, one has $f_n \to f$ in W_p^r if $\|f_n^{(r)} - f^{(r)}\|_p(A) \to 0$ and if for some $a \in A$, $f_n^{(k)}(a) \to f^{(k)}(a)$, $k = 0,1,\dots,r-1$; for $A = \mathbb{T}$, it is enough to assume this for $k = 0$.

Indeed, if functions g_n, $n = 1,2,\dots$, are absolutely continuous, (a) $g_n' \to 0$ in $L_p(A)$ and (b) $g_n(a) \to 0$, then $g_n(x) \to 0$ uniformly, for $g_n(x) = g_n(a) + \int_a^x g_n'\,dt \to 0$. Also, if (a) $g_n' \to 0$ in $L_p(\mathbb{T})$ and (b) the g_n themselves are derivatives, then they have mean values zero, hence $g_n(a_n) = 0$ for some $a_n \in \mathbb{T}$, $n = 1,2,\dots$, and again $g_n(x) = \int_{a_n}^x g_n'\,dt \to 0$ uniformly.

We can prove that the $W_p^r(A)$ are Banach spaces. Let (f_n) be a Cauchy sequence in W_p^r. Then, $f_n^{(k)}$, $k = 0,\dots,r$, are Cauchy sequences in L_p and hence converge: $f_n^{(k)} \to g_k$, $n \to \infty$, in $L_p(A)$. From (5.15), we see that $f_n^{(k)}$ converges uniformly to g_k on compact subsets of A for all $k < r$. Hence for $x, y \in A$ and $k < r$,

$$g_k(x) = g_k(y) + \lim_{n\to\infty} \int_y^x f_n^{(k+1)}\, dt = g_k(y) + \int_y^x g_{k+1}\, dt\,.$$

Thus $g_k = g_0^{(k)}$, $k = 0, \dots, r$, $g_0 \in W_p^r$ and $\|f_n - g_0\|_{W_p^r} \to 0$.

Proposition 5.8. *For each of the four possible A, and for $r = 1, 2, \dots$, the space $C^\infty(A)$ is dense in $W_p^r(A)$, $1 \le p < \infty$, and in $C^r(A)$, $p = \infty$.*

Proof. First let A be compact. Then each $f \in W_p^r(A)$ can be approximated by polynomials in the case $A = [a, b]$, by trigonometric polynomials if $A = \mathbb{T}$. Consider for instance the last case. We take trigonometric polynomials S_n such that $\|f^{(r)} - S_n\|_p \to 0$, $n \to \infty$, and let T_n be their r-th periodic integrals normalized by $T_n(0) = f(0)$. By previous remarks, $T_n \to f$ in $W_p^r(\mathbb{T})$.

Now let A be infinite, for example $A = \mathbb{R}$. By Theorem 5.2, there is a partition of unity $1 = \sum_{j\in\mathbb{Z}} \phi_j$, which consists of functions $\phi_j \in C^\infty$ with supports $I_j := [j-1, j+1]$ satisfying $\|\phi_j\|_{W_\infty^r} \le M$ with M depending only on r. If $\varepsilon > 0$, we select for each $j \in \mathbb{Z}$ a function $f_j \in C^\infty(I_j)$ with $\|f - f_j\|_{W_p^r(I_j)} < 2^{-|j|}\varepsilon$ and extend f_j by zero outside of I_j. We then have $\phi_j f_j \in C^\infty(\mathbb{R})$ and by Leibniz' rule, Theorem 5.2, (iii) and by (5.8),

$$\|(f - f_j)\phi_j\|_{W_p^r(\mathbb{R})} \le C\|f - f_j\|_{W_p^r(I_j)} \le C2^{-|j|}\varepsilon\,.$$

Now, $h := \sum_{j\in\mathbb{Z}} \phi_j f_j$ is in $C^\infty(\mathbb{R})$ and $f - h = \sum_{j\in\mathbb{Z}}(f - f_j)\phi_j$. We deduce

$$\|f - h\|_{W_p^r(\mathbb{R})} \le C\varepsilon \sum_{j\in\mathbb{Z}} 2^{-|j|} < 3C\varepsilon\,. \qquad \square$$

The space W_1^1 consists of functions f with an integrable derivative; in this case $\mathrm{Var}_A f = \int_A |f'|\, dx$. The derivatives of functions $f \in BV$ are integrable, but we do not have the representation $f(x) = f(c) + \int_c^x f'(t)\, dt$. Thus, $W_1^1(A)$ is isometric to a nontrivial subspace of $BV(A)$, and even of $BV(A) \cap C(A)$.

§ 6. Moduli of Continuity

Measuring the smoothness of a function by differentiability is too crude for many purposes in approximation. More subtle measurements are provided by the moduli of continuity and the moduli of smoothness of §7.

The *modulus of continuity* $\omega(f, t) =: \omega(t)$ of a function f can be defined when f is given on any metric space A, but we shall restrict ourselves to $A = \mathbb{R}$, $\mathbb{R}_+$, $[a, b]$, or $\mathbb{T}$. In that case

$$(6.1) \qquad \omega(t) := \omega(f, t) := \sup_{\substack{|x-y|\le t\\ x,y\in A}} |f(x) - f(y)|\,, \qquad t \ge 0\,.$$

Clearly, $\omega(t)$ is constant for $t \ge \operatorname{diam} A$, if A is bounded. The function ω is continuous at $t = 0$ if and only if f is uniformly continuous on A. We shall assume that $f \in \widetilde{C}(A)$ – that f belongs to the space of uniformly continuous functions on A. Then $\omega(f, t)$ is finite for each t. For each fixed t, ω is a semi-norm, that is, it is subadditive in f and positive homogenous.

A modulus of continuity has the following simple properties:

$$(6.2) \qquad \begin{array}{ll} \text{(a)} & \omega(t) \to \omega(0) = 0\,, \quad \text{for } t \to 0\,; \\ \text{(b)} & \omega \text{ is non-negative and non-decreasing on } \mathbb{R}_+\,; \\ \text{(c)} & \omega \text{ is } \textit{subadditive}\text{: } \omega(t_1 + t_2) \le \omega(t_1) + \omega(t_2)\,; \\ \text{(d)} & \omega \text{ is continuous on } \mathbb{R}_+\,. \end{array}$$

Properties (a) and (b) are clear. For (c), if $|x - y| \le t_1 + t_2$, there is a point $z \in A$ for which $|x - z| \le t_1$, $|y - z| \le t_2$, and (c) follows from

$$|f(x) - f(y)| \le |f(x) - f(z)| + |f(z) - f(y)| \le \omega(t_1) + \omega(t_2)\,.$$

Moreover,

$$(6.3) \qquad |\omega(t_1 + t_2) - \omega(t_1)| \le \omega(t_2)\,.$$

Thus, (a), (b), (c) imply that ω is continuous at each $t \ge 0$.

A function ω defined on $\mathbb{R}_+$ and satisfying (6.2) is called *a modulus of continuity*. This is justified since by (6.3), any such function is its own modulus of continuity.

It follows from (6.2)(c) by induction that

$$\omega(t_1 + \dots + t_n) \le \omega(t_1) + \dots + \omega(t_n)\,.$$

For $t = t_1 = \dots = t_n$, we obtain

$$(6.4) \qquad \omega(nt) \le n\omega(t)\,.$$

A similar inequality holds for a nonintegral factor λ:

$$(6.5) \qquad \omega(\lambda t) \le (\lambda + 1)\omega(t)\,, \qquad \lambda \ge 0\,.$$

In fact, taking an integer n for which $n \le \lambda < n + 1$, we see that

$$\omega(\lambda t) \le \omega\big((n+1)t\big) \le (n+1)\omega(t) \le (\lambda + 1)\omega(t)\,.$$

A modulus of continuity cannot be too small. If $\omega(f, t)/t \to 0$ for $t \to 0$, then $f'(x) \equiv 0$ and f is a constant.

For a concave function f on $[a, b]$, $\alpha f(x) + \beta f(y) \le f(\alpha x + \beta y)$ for $x, y \in [a, b]$ and $\alpha \ge 0$, $\beta \ge 0$, $\alpha + \beta = 1$. A concave function f on $[0, 1]$, which satisfies $f(0) = 0$, has the property that $f(x)/x$ decreases, for if $x < y$, then

$$\frac{x}{y} f(y) = \frac{y - x}{y} f(0) + \frac{x}{y} f(y) \le f(x)\,.$$

Example 1. A continuous, increasing function ω on $\mathbb{R}_+$, which satisfies $\omega(0) = 0$, is a modulus of continuity if it is concave (or, more generally, if $\omega(t)/t$

is decreasing). It is necessary only to show that ω satisfies (6.2)(c). This is obtained by multiplying the inequalities

$$\frac{\omega(t_1+t_2)}{t_1+t_2} \le \frac{\omega(t_1)}{t_1} \quad \text{and} \quad \frac{\omega(t_1+t_2)}{t_1+t_2} \le \frac{\omega(t_2)}{t_2}$$

by t_1 and t_2, respectively, and adding.

Example 2. We shall need the modulus of continuity of the function F below. Let ω be a modulus of continuity and let $I_i := (\alpha_i, \alpha_i + 2\delta)$, $\delta > 0$, be finitely many disjoint intervals contained in $[a,b]$. Let F be zero outside the I_i, and let

$$F(x) := \begin{cases} c_i\omega(x-\alpha_i), & \alpha_i \le x \le \alpha_i + \delta, \\ c_i\omega(\alpha_i + 2\delta - x), & \alpha_i + \delta \le x \le \alpha_i + 2\delta. \end{cases} \tag{6.6}$$

If there is only one interval I_1 with $c_1 = c > 0$, and if $t \le \delta$, then from (6.3) the maximum of $|F(x+h) - F(x)|$ for $|h| \le t$ is attained when $x = \alpha_1$, $h = t$, and is equal to $c\omega(t)$. It is $c\omega(\delta)$ if $t \ge \delta$. Hence,

$$\omega(F,t) = \begin{cases} c\omega(t), & 0 \le t \le \delta, \\ c\omega(\delta), & t \ge \delta. \end{cases} \tag{6.7}$$

This remains true for several I_i if all $c_i > 0$ and $c = \max c_i$. If the c_i change sign, let $c = \max |c_i|$. Then F is the difference of two functions of type (6.6) with positive c_i, and we have (6.7) with c replaced by $2c$ and equality replaced by inequality.

Example 3. If ω is concave, a better estimate holds. Assuming, as before, that $c = \max |c_i|$, and that F is given by (6.6), we have

$$\omega(F,t) \le \begin{cases} 2c\omega(\frac{1}{2}t), & 0 \le t \le 2\delta, \\ 2c\omega(\delta), & t \ge 2\delta. \end{cases} \tag{6.8}$$

In fact, the largest possible value of $|F(x+t) - F(x)|$, for a fixed t with $0 < t \le 2\delta$, happens when $x \in I_{i-1}$, $(x+t) \in I_i$, with the intervals I_{i-1}, I_i having the common endpoint α_i, and when $c_{i-1} = -c_i = \pm c$. If $x = \alpha_i - t_1$, $0 \le t_1 \le t$, then by the concavity of ω,

$$|F(x+t) - F(x)| \le c\omega(t_1) + c\omega(t - t_1) \le 2c\omega(\frac{1}{2}t).$$

If the modulus of continuity ω is not concave, it can be often replaced by its concave majorant. We first make the following useful observation: if L is any collection of linear (more generally concave) functions, then, assuming that the function on $\mathbb{R}_+$

$$\psi(t) := \inf_{l \in L} l(t), \tag{6.9}$$

is finite, it is concave. This can be used as follows. Let ϕ be a function on $\mathbb{R}_+$, and let L be the collection of all linear functions l for which $l(t) \ge \phi(t)$, $t \in \mathbb{R}_+$; we assume that such l exist. From (6.9) we see that

$$\overline{\phi}(t) := \inf_{l \in L} l(t), \qquad t > 0 \tag{6.10}$$

is a concave function with $\overline{\phi} \ge \phi$. Moreover, if ψ is concave and $\psi(t) \ge \phi(t)$ for $t > 0$, then also $\psi(t) \ge \overline{\phi}(t)$, $t > 0$. To prove this, we use the fact that at each point $t_0 > 0$, there exist finite right and left derivatives $\psi'_+(t_0)$, $\psi'_-(t_0)$ and $\psi'_+(t_0) \le \psi'_-(t_0)$. If M lies between these two numbers, and l is the linear function with slope M which interpolates ψ at t_0, then l is a *supporting linear function*: it satisfies $l(t) \ge \psi(t)$ for all t and $l(t_0) = \psi(t_0)$. In application to $\overline{\phi}$, we get the desired inequality $\overline{\phi}(t_0) \le \psi(t_0)$. This inequality shows that the function (6.10) is the *least concave majorant* of ϕ. If ϕ is a bounded function with $\phi(0) = \phi(0+) = 0$, then $\overline{\phi}$ also has these properties.

Lemma 6.1. *If ω is a modulus of continuity then its least concave majorant $\overline{\omega}$ is also a modulus of continuity, and satisfies*

$$\overline{\omega}(t) \le 2\omega(t), \qquad t > 0. \tag{6.11}$$

Proof. Let $t_0 \ge 0$. We define

$$l(t) = 2\omega(t_0) + \frac{\omega(t_0)}{t_0}(t - t_0).$$

Then $\omega(t) \le \omega(t_0) \le l(t)$ for $0 \le t \le t_0$. This is also true for $t \ge t_0$, for in this case by (6.5)

$$\omega(t) \le \left(\frac{t}{t_0} + 1\right)\omega(t_0) = l(t).$$

Thus, for the least concave majorant $\overline{\omega}$ of ω we have $\overline{\omega}(t_0) \le l(t_0) = 2\omega(t_0)$. Since $t_0 > 0$ is arbitrary, (6.11) holds for all $t > 0$, and by continuity also for $t = 0$. □

Example 4. For $0 < a < 1$, the function ω on $[0,2]$ given by $\omega(t) := t/a$, $0 \le t \le a$, $\omega(t) := 1$, $a \le t \le 1$, $\omega(t) := \omega(t-1) + 1$, $1 \le t \le 2$ satisfies (6.2) and is an example of a non-concave modulus of continuity. Since $\overline{\omega}(1) = (2-a)\omega(1)$, we see that 2 is the best possible constant in (6.11).

One can define moduli of continuity $\omega(f,t)_X$ for any rearrangement-invariant space X, for example for $L_p(A)$, $0 < p < \infty$. We shall do this assuming that $A = \mathbb{T}$, $\mathbb{R}$, $\mathbb{R}_+$, or $[a,b]$. A useful notation is $A_h := [a, b-h]$ if $A = [a,b]$, $0 < h < b - a$ and $A_h := \emptyset$, $h \ge b - a$. For other possible A, we define $A_h := A$, $h > 0$. We let $T_h, h \in \mathbb{R}$, denote the translation operator $T_h(f,x) := f(x+h)$ and $\Delta_h := T_h - I$ the difference operator where I is the identity operator. The modulus of continuity for $X = L_p$ is then

$$\omega(f,t,A)_p := \sup_{0\le h\le t} \|\Delta_h(f)\|_p(A_h)\,. \tag{6.12}$$

It follows that $\omega(f,t)_p$ has properties (6.2) (a), (b), (d) for $f \in L_p$. If $p \ge 1$ then (c) also is valid; hence $\omega(f,t)_p$ is a modulus of continuity. If $p=\infty$ and $f \in C$, (6.12) reduces to (6.1).

§ 7. Moduli of Smoothness

If a continuous function f on an interval A satisfies $\omega(f,t)=o(t)$, then f is constant. Thus the modulus of continuity is not useful for measuring higher smoothness. For the latter, we shall use moduli of smoothness which are connected with differences of higher orders. In this section, as before, A is $\mathbb{T}$, or a subinterval of $\mathbb{R}$.

From the first difference $\Delta_h := T_h - I$, we define by induction on r the higher differences $\Delta_h^r := \Delta_h[\Delta_h^{r-1}] = (T_h - I)^r$, $r=1,2,\dots, h\in\mathbb{R}$. It follows from the binomial theorem that

$$\Delta_h^r(f,x) = \sum_{k=0}^{r} \binom{r}{k} (-1)^{r-k} f(x+kh)\,. \tag{7.1}$$

Also, we have $\Delta_{-h}^r(f,x) = (-1)^r \Delta_h^r(f, x-rh)$. For a function f on A, and $h \ge 0$, $\Delta_h^r(f,x)$ is defined for all $x \in A_{rh}$. The set A_{rh} coincides with A except when $A=[a,b]$, in which case $A_{rh} := [a, b-rh]$.

The r*-th modulus of smoothness* of $f \in L_p(A)$, $0<p<\infty$, or of $f\in C(A)$, for compact A, of $f \in \widetilde{C}(A)$ for non-compact A, if $p=\infty$, is defined by

$$\omega_r(f,t)_p := \sup_{0<h\le t} \|\Delta_h^r(f,\cdot)\|_p(A_{rh})\,, \qquad t\ge 0\,. \tag{7.2}$$

Thus, $\omega_1(f,t)_p = \omega(f,t)_p$ is the modulus of continuity. We do not indicate the dependence of ω_r on p, if there is no possibility of confusion.

It is clear that $\omega_r(f,t)_p$ is finite for each t; it is a continuous and increasing function of t with $\omega_r(f,0)_p = 0$. For $1\le p\le\infty$, the triangular inequality shows that

$$\omega_r(f+g,t)_p \le \omega_r(f,t)_p + \omega_r(g,t)_p\,, \qquad f,g\in L_p\,. \tag{7.3}$$

Also, $\omega_r(cf,t)_p = |c|\omega_r(f,t)_p$. Therefore, $\omega_r(\cdot,t)_p$ is a seminorm on L_p; it vanishes for all polynomials in $\mathcal{P}_{r-1}$ (or for all constants if $A=\mathbb{T}$, or if A is not compact). When $p<1$, (7.3) is replaced by

$$\omega_r(f+g,t)_p^p \le \omega_r(f,t)_p^p + \omega_r(g,t)_p^p\,. \tag{7.4}$$

From (7.1), we see that the norm of the operator Δ_h^r in $L_p(A)$, $1\le p\le\infty$, does not exceed 2^r. We define $\omega_0(f,t)_p = \|f\|_p$ and have for $0\le k\le r$

$$\begin{cases} \omega_r(f,t)_p \le 2^{r-k}\omega_k(f,t)_p\,, & 1\le p\le\infty\,,\\ \omega_r(f,t)_p^p \le 2^{r-k}\omega_k(f,t)_p^p\,, & 0<p<1\,. \end{cases} \tag{7.5}$$

There is also a formula similar to (6.4). It follows from the relation

$$\Delta_{nh}^r(f,x) = \sum_{k_1=0}^{n-1} \cdots \sum_{k_r=0}^{n-1} \Delta_h^r(f, x+k_1h+\cdots+k_rh)\,. \tag{7.6}$$

If $r=1$, this is obvious; for arbitrary r, it is proved by induction by applying the operator Δ_{nh} to both sides of (7.6). From (7.6) we obtain for $1\le p\le\infty$:

$$\omega_r(f,nt)_p \le n^r\omega_r(f,t)_p\,. \tag{7.7}$$

For $p<1$ the corresponding inequality is $\omega_r(f,nt)_p^p \le n^r\omega_r(f,t)_p^p$. As in (6.5), from (7.7) one derives

$$\omega_r(f,\lambda t)_p \le (\lambda+1)^r\omega_r(f,t)_p\,, \qquad \lambda>0\,. \tag{7.8}$$

There is a useful integral representation of $\Delta_h^r f$ by means of a derivative of f. Let $M_1(x) := \chi_{[0,1]}$ where χ_B denotes the characteristic function of the set B. We let M_r be the r-fold convolution of M_1 with itself; then $M_r(x) := \int_{\mathbb{R}} M_{r-1}(x-y)M_1(y)\,dy$. By induction, it follows that M_r is supported on $[0,r]$, has integral one and that $0\le M_r\le 1$. In Chapter 5, we see that M_r is a B-spline – a piecewise polynomial with breakpoints $0,1,\dots,r$. For $h>0$, the function $M_r(x,h) := h^{-1}M_r(h^{-1}x)$ has analogous properties.

If f' is locally integrable, then

$$h^{-1}\Delta_h f(x) = h^{-1}\int_x^{x+h} f'(t)\,dt = \int_{\mathbb{R}} f'(x+t)M_1(t;h)\,dt\,. \tag{7.9}$$

More generally, we have

$$\begin{aligned} h^{-r}\Delta_h^r(f,x) &= \int_{\mathbb{R}} f^{(r)}(x+t)M_r(t;h)\,dt \\ &= \int_{\mathbb{R}} f^{(r)}(t)M_r(t-x;h)\,dt\,, \qquad x\in A_{rh}\,, \end{aligned} \tag{7.10}$$

whenever $f,\dots,f^{(r-1)}$ are locally absolutely continuous. Indeed, using (7.9) and induction on r, we find

$$\begin{aligned} h^{-r}\Delta_h^r(f,x) &= h^{-r}\Delta_h^{r-1}(\Delta_h(f),x) \\ &= \int_0^{(r-1)h} du \int_0^h f^{(r)}(x+u+v)M_{r-1}(u)M_1(v)\,dv \end{aligned}$$

and (7.10) follows by making the change of variables $u=t-v$. For more general formulas than (7.10), see (7.16) of Chapter 4. The kernel $M_r(t-x;h)$ in (7.10) is usually called the Peano kernel of Δ_h^r.

Since $\int_A M_r(t;h)\,dt = 1$, we can use a continuous analogue of Theorem 2.1 of Chapter 1 (with the variable $n\to\infty$ replaced by $h\to 0+$) and obtain

$$\lim_{h\to 0} h^{-r}\Delta_h^r(f,x) = f^{(r)}(x) \tag{7.11}$$

locally uniformly if $f \in C^r(A)$. Similarly, in Theorem 2.4 of Chapter 1 we can take the bell-shaped majorant $L_h(x,t) := h^{-1}$ on $[x-rh, x+rh]$, $:= 0$ elsewhere and from $\int_{\mathbb{R}} L_h\,dt \le 2r$ deduce (7.11) *almost everywhere* for all $f \in W_p^r(A)$.

By means of the Minkowski's inequality (1.6), we derive from (7.10) that for $1 \le p \le \infty$,

$$\omega_r(f,t)_p \le t^r |f|_{W_p^r}, \qquad t \ge 0. \tag{7.12}$$

On the other hand, if we first apply Δ_h^k to both sides of (7.10) and then use Minkowski's inequality, we obtain

$$\omega_{r+k}(f,t)_p \le t^r \omega_k(f^{(r)},t)_p, \qquad t \ge 0. \tag{7.13}$$

Proposition 7.1. *If $A = \mathbb{R}$, $\mathbb{R}_+$, $[a,b]$, or $\mathbb{T}$, then for $f \in L_p(A)$, $1 \le p < \infty$, and for uniformly continuous functions $f \in \widetilde{C}(A)$ for $p = \infty$,* (i) *the modulus of smoothness $\omega_r(f,t)_p$ vanishes identically if it is zero for some $t > 0$;* (ii) *if $\|\Delta_h^r(f,\cdot)\|_p = o(h^r)$, $h \to 0+$, then f is a.e. a polynomial $P \in \mathcal{P}_{r-1}$;* (iii) *if $f \notin \mathcal{P}_{r-1}$, one has $\omega_r(f,t)_p \ge C\,t^r$, $C > 0$, $0 < t \le 1$.*

Proof. (i) follows directly from (7.8). (ii) Let I be any subinterval of A which is contained in A_{rh} for all sufficiently small h. If $\varphi \in C_0^\infty(I)$, then by Hölder's inequality, for $h > 0$,

$$\left|\int_{\mathbb{R}} f(x) h^{-r}\Delta_{-h}^r(\varphi,x)\,dx\right| = \left|\int_{\mathbb{R}} h^{-r}\Delta_h^r(f,x)\varphi(x)\,dx\right| \le \varepsilon(h)\|\varphi\|_{p'}. \tag{7.14}$$

We let $h \to 0^+$ in (7.14) and use (7.11) to find

$$\int_{\mathbb{R}} f(x)\varphi^{(r)}(x)\,dx = 0.$$

Hence the generalized r-th derivative of f is 0 on I and Theorem 5.3 gives that $f = P_I$ a.e. on I for some polynomial $P_I \in \mathcal{P}_{r-1}$. By taking intervals which overlap, we see that all of the P_I are identical. (iii) For $\lambda = 1/t$ one gets from (7.8), (i) and (ii) that $\omega_r(f,t)_p \ge 2^{-r}\omega_r(f,1)_p t^r = C\,t^r$, $C > 0$ for all $0 < t \le 1$. □

(For additional information for the case $0 < p < 1$ see §5 of Chapter 12.)

§ 8. Marchaud Inequalities

The inequality $\omega_r(f,t)_p \le 2^{r-k}\omega_k(f,t)_p$, $1 \le k < r$, $1 \le p \le +\infty$ of (7.5) is an estimate of ω_r by means of smoothness moduli ω_k of lower order. Inverse estimates are less simple. If f is a polynomial of degree k, then $\omega_k(f,t) = ct^k$,

$c > 0$, while $\omega_r(f,t) \equiv 0$; so we cannot simply reverse the above estimate. Even if we assume that $\omega_r(f,t) \ne 0$, the example of $f = \varepsilon t^r + t^k$ shows that $\omega_r(f,t)$ can be very small, while $\omega_k(f,t)$ is again of order of ct^k.

Marchaud [1927] has shown that (7.5) has a weak inverse: one can estimate ω_k by means of an integral of ω_r:

Theorem 8.1. *There is a constant C which depends on $r = 2,3,\dots$ with the following property. If $f \in L_p(A)$, $1 \le p < \infty$, or $f \in C(A)$, $p = \infty$, and if $1 \le k < r$, then*

$$\omega_k(f,t)_p \le Ct^k \int_t^\infty \frac{\omega_r(f,s)_p\,ds}{s^{k+1}}, \qquad t > 0, \tag{8.1}$$

except when $A = [a,b]$, in which case

$$\omega_k(f,t)_p \le Ct^k\left\{\int_t^{|A|} \frac{\omega_r(f,s)_p\,ds}{s^{k+1}} + \frac{\|f\|_p}{|A|^k}\right\}, \qquad t > 0. \tag{8.2}$$

Proof. We shall suppress the subscript p. We first prove the theorem when $r = k+1$ and later derive from this the general case. For the polynomial $Q(x) := [1 - 2^{-k}(x+1)^k]/(x-1)$ of degree $k-1$ we have $(x-1)^k = 2^{-k}(x^2-1)^k + Q(x)(x-1)^{k+1}$. Replacing x by the translation operator T_h, we obtain

$$(T_h - I)^k = 2^{-k}(T_{2h} - I)^k + Q(T_h)(T_h - I)^{k+1}. \tag{8.3}$$

Here and later, we use the commutivity of the translation operators: $T_sT_t = T_tT_s$ for all s and t.

Now let $A' := [a,c]$, $c := (a+b)/2$ if $A = [a,b]$ and $A' := A$ in all other cases. The operators T_h^j all have norm one on L_p; therefore $M := \|Q(T_h)\|_p$ does not exceed the sum of the absolute values of the coefficients of Q. From (8.3) we derive

$$\begin{aligned}\|\Delta_h^k(f)\|(A') &\le 2^{-k}\|\Delta_{2h}^k(f)\|(A') + M\|\Delta_h^{k+1}(f)\|(A') \\ &\le 2^{-k}\|\Delta_{2h}^k(f)\|(A') + M\omega_{k+1}(f,h),\end{aligned} \tag{8.4}$$

provided that $4kh \le |A|$ in the case $A = [a,b]$. We repeatedly apply (8.4) and obtain

$$\|\Delta_h^k(f)\|(A') \le M\sum_{j=0}^m 2^{-kj}\omega_{k+1}(f,2^jh) + 2^{-km}\|f\|(A) \tag{8.5}$$

provided that $2^{m+2}kh \le |A|$ in the case $A = [a,b]$.

Excluding for the moment the case $A = [a,b]$, we let $m \to \infty$ in (8.5) and then take the supremum over $0 \le h \le t$ to obtain

$$\omega_k(f,t) \le Mt^k\sum_{j=0}^\infty (2^jt)^{-k}\omega_{k+1}(f,2^jt). \tag{8.6}$$

This implies (8.1) for $r = k+1$, since the j-th term of the last sum does not exceed
$$2^k \int_{2^j t}^{2^{j+1}t} \omega_{k+1}(f,s)s^{-k-1}\,ds\,.$$

If $A = [a,b]$, then (8.5) holds also with A' replaced by $A'' := [c,b]$. This can be obtained by applying (8.5) to the function $g(x) := f(b-x)$ which has the same moduli of smoothness as f (because $\Delta_h^s(f,x) = (-1)^s \Delta_h^s(g, b-x-sh)$, $s = 1,2,\dots$). It follows that (8.5) also holds with A in place of A' and with an additional factor $2^{1/p}$ on the right. Therefore,

$$\omega_k(f,t) \le 2^{1/p} M t^k \sum_{j=0}^{m} 2^{-kj}\omega_{k+1}(f, 2^j t) + 2^{-km}2^{1/p}\|f\|\,. \tag{8.7}$$

We take m to be the first integer for which $2^{m-1}t \ge |A|$, then (8.2) follows from (8.7).

Finally, we can prove (8.1) and (8.2) for arbitrary k, r by induction on r. We know that these inequalities are valid for $r = k+1$. It will be sufficient to show that if they hold for a pair k, r, $r \ge k+1$, then also for the pair $k, r+1$. Under these conditions, from (8.1) we obtain

$$\begin{aligned}
\omega_k(f,h) &\le Ch^k \int_h^\infty s^{-k-1}\omega_r(f,s)\,ds \\
&\le Ch^k \int_h^\infty s^{r-k-1} ds \int_s^\infty t^{-r-1}\omega_{r+1}(f,t)\,dt \\
&= Ch^k \int_h^\infty t^{-r-1}\omega_{r+1}(f,t)\,dt \int_h^t s^{r-k-1}\,ds \\
&\le Ch^k \int_h^\infty t^{-k-1}\omega_{r+1}(f,t)\,dt\,,
\end{aligned} \tag{8.8}$$

as required. A similar argument applies to (8.2). □

There are also Marchaud inequalities for $p < 1$, which will be needed in Chapter 12.

Theorem 8.2. *Let $0 < p < 1$ and $1 \le k < r$. Then there is a constant C depending only on r and p such that for each $f \in L_p$,*

$$\omega_k(f,t)_p^p \le Ct^{kp} \int_t^\infty \frac{\omega_r(f,s)_p^p\,ds}{s^{kp+1}}\,, \qquad t > 0\,, \tag{8.9}$$

except when $A = [a,b]$, in which case

$$\omega_k(f,t)_p^p \le Ct^{kp}\left\{\int_t^{|A|} \frac{\omega_r(f,s)_p^p\,ds}{s^{kp+1}} + \frac{\|f\|_p^p}{|A|^{kp}}\right\}. \tag{8.10}$$

Proof. The proof is similar to that of Theorem 8.1. We indicate the changes needed in the proof of (8.1) in order to arrive at (8.9). Similar changes in the proof of (8.2) give (8.10). We use the subadditivity of $\|\cdot\|_p^p$; this introduces a power of p in (8.4)–(8.6) resulting in the inequality

$$\omega_k(f,t)_p^p \le Ct^{kp} \int_t^\infty \frac{\omega_{k+1}(f,s)_p^p\,ds}{s^{kp+1}}\,, \qquad t > 0\,.$$

This is (8.9) when $r = k+1$. Assuming that (8.9) is valid for a pair k, r, with $r \ge k+1$, by means of estimates used in (8.8), we establish its truth for k, $r+1$. Thus (8.9) follows by induction. □

There is an improvement of Theorem 8.1 due to M.F. Timan [1958] which follows from elementary properties of L_p norms. Here and later we use ideas of Ditzian [1988]. If $0 < p \le \infty$, let $\mu := \min(p,2)$. Then for any $a, b \in \mathbb{R}$, and $M = M(p)$ sufficiently large, we have

$$|a+b|^p + |a-b|^p \le 2(|a|^\mu + M|b|^\mu)^{p/\mu}\,, \qquad 0 < p < \infty\,. \tag{8.11}$$

By homogeneity, it is sufficient to consider $a = 1$ and $b = x > 0$ and prove that
$$f(x) := |1+x|^p + |1-x|^p \le 2(1 + Mx^\mu)^{p/\mu} =: g(x)\,.$$

This inequality is clear (with constant M depending on p and x_0) if $x \ge x_0$ and $x_0 > 0$ is fixed. Therefore, it is enough to show that $f(x) \le g(x)$ holds in a neighborhood of 0. This is trivial for $p = 1$. For $0 < p < 1$, this follows because $f(0) = g(0) = 2$, and $f'(x) \le g'(x)$ for x close to 0. If $1 < p < \infty$, we have $f(0) = g(0) = 2$, $f'(0) = g'(0) = 0$, and $f''(x) \le g''(x)$ if x is close to 0.

Lemma 8.3. *Let $1 \le p < \infty$ and $\mu := \min(p,2)$. For the constant $M := M(p)$ of (8.11), for all $f, g \in L_p$ we have*

$$\|f+g\|_p + \|f-g\|_p \le 2(\|f\|_p^\mu + M\|g\|_p^\mu)^{1/\mu}\,. \tag{8.12}$$

Proof. We take $a := f(x)$, $b := g(x)$ in (8.11) and integrate to obtain

$$\|f+g\|_p^p + \|f-g\|_p^p \le 2\int_A (|f|^\mu + M|g|^\mu)^{p/\mu}\,dx\,. \tag{8.13}$$

Now, the left side of (8.12) does not exceed $2^{1-1/p}(\|f+g\|_p^p + \|f-g\|_p^p)^{1/p}$ while the right side of (8.13) by Minkowski's inequality does not exceed $2(\|f\|_p^\mu + M\|g\|_p^\mu)^{p/\mu}$. □

Theorem 8.4 (Timan [1958]). *There is a constant C which depends on $r = 2, 3, \dots$ with the following property. If $f \in L_p(A)$, $1 \le p < \infty$, and if $1 \le k < r$, then*

$$\omega_k(f,t)_p \le Ct^k\left(\int_t^\infty \frac{\omega_r(f,s)_p^\mu\,ds}{s^{k\mu+1}}\right)^{1/\mu}, \qquad t > 0\,, \tag{8.14}$$

except when $A = [a,b]$, in which case

$$\omega_k(f,t)_p \le Ct^k\left(\int_t^{|A|}\frac{\omega_r(f,s)_p^\mu\,ds}{s^{k\mu+1}}+\frac{\|f\|_p^\mu}{|A|^{k\mu}}\right)^{1/\mu},\qquad t>0. \tag{8.15}$$

Proof. The proof is similar to that of Theorem 8.1. We consider the case $A=\mathbb{R}$ or $\mathbb{T}$ and suppress the subscript p; the modification to handle the case $A=[a,b]$ is the same as in Theorem 8.1. Let $T:=T_t$ again be the translation operator. We shall use the operators $S_1:=\frac{1}{2}(T^2-I)$, $S_2:=\frac{1}{2}(T-I)^2$, which satisfy $T-I=S_1-S_2$, $T(T-I)=S_1+S_2$. Instead of (8.3) we can use the two identities

$$\begin{aligned}(T-I)^r&=(T-I)^{r-1}S_1-(T-I)^{r-1}S_2\,,\\ T(T-I)^r&=(T-I)^{r-1}S_1+(T-I)^{r-1}S_2\,.\end{aligned}\tag{8.16}$$

For $f\in L_p$, let $f_i:=(T-I)^{r-1}S_if$, $i=1,2$. Then, using Lemma 8.3 and the fact that T is an isometry,

$$\begin{aligned}\|(T-I)^rf\|&=\tfrac{1}{2}\|(T-I)^rf\|+\tfrac{1}{2}\|T(T-I)^rf\|\\&=\tfrac{1}{2}\|f_1-f_2\|+\tfrac{1}{2}\|f_1+f_2\|\le\{\|f_1\|^\mu+M\|f_2\|^\mu\}^{1/\mu}\\&=\tfrac{1}{2}\{\|(T-I)^{r-1}(T^2-I)f\|^\mu+M\|(T-I)^{r+1}f\|^\mu\}^{1/\mu}.\end{aligned}$$

Replacing here r by $r-i$, f by $2^{-i}(T^2-I)^if$, we see that

$$\begin{aligned}2^{-i\mu}\|(T-I)^{r-i}(T^2-I)^if\|^\mu\le 2^{-(i+1)\mu}\{&\|(T-I)^{r-i-1}(T^2-I)^{i+1}f\|^\mu\\&+M\|(T-I)^{r-i+1}(T^2-I)^if\|^\mu\}.\end{aligned}$$

Repeating this operation, we obtain for $(T-I)^rf$,

$$\begin{aligned}\|(T-I)^rf\|^\mu&\le 2^{-r\mu}\{\|(T^2-I)^rf\|^\mu\\&\qquad+M\sum_{j=0}^{r-1}2^{-(j+1)\mu}\|(T-I)^{r-j+1}(T^2-I)^jf\|^\mu\}\\&\le 2^{-r\mu}\{\|(T^2-I)^rf\|^\mu+M_1\|(T-I)^{r+1}f\|^\mu\},\end{aligned}$$

because each of the terms in the sum has the factor $(T-I)^{r+1}f$. It follows that

$$\omega_r(f,t)_p^\mu\le 2^{-r\mu}\omega_r(f,2t)_p^\mu+M_1\omega_{r+1}(f,t)_p^\mu\,.$$

Iterating this inequality as in the proof of Theorem 8.1, we derive (8.14) for $k=r+1$. For general k, we obtain (8.14) in a similar manner to (8.8). □

Remark. That (8.14) is better than (8.1) can be seen by discretizing both integrals. From the monotonicity and subadditivity of ω_r, it follows that the integral on the right of (8.14) does not exceed

$$C\left(\sum_{2^j\ge t}2^{-jk\mu}\omega_r(f,2^j)_p^\mu\right)^{1/\mu},$$

while the integral on the right of (8.1) is

$$\ge C\sum_{2^j\ge t}2^{-jk}\omega_r(f,2^j)_p\,.$$

Since an l_μ norm does not exceed an l_1 norm (because $\mu\ge1$), the right side of (8.14) does not exceed a multiple of the right side of (8.1).

§ 9. Lipschitz Spaces

The simplest Lipschitz space $\operatorname{Lip}\alpha$, $0<\alpha\le1$, consists of all continuous functions $f\in\widetilde{C}(A)$ on a set A which satisfy

$$|\Delta_t(f,x)|=|f(x+t)-f(x)|\le Mt^\alpha,\qquad t>0, \tag{9.1}$$

or, equivalently, $\omega(f,t)\le Mt^\alpha$. In the latter form one can define $\operatorname{Lip}\alpha$ on any metric space. We shall restrict ourselves to $A=[a,b]$, $\mathbb{T}$, $\mathbb{R}_+$, $\mathbb{R}$. The semi-norm of $\operatorname{Lip}\alpha$ is

$$|f|_{\operatorname{Lip}\alpha}:=\sup_{t>0}\big(t^{-\alpha}\omega(f,t)\big)\,. \tag{9.2}$$

Taking the norm of $\Delta_t(f,x)$ in some space X, we obtain spaces $\operatorname{Lip}(\alpha,X)$, for example, for the L_p-norm, the space $\operatorname{Lip}(\alpha,L_p)$ which consists of all $f\in L_p$, $0<p\le\infty$ for which

$$\|\Delta_t(f,x)\|_p=\left\{\int_{A_t}|f(x+t)-f(x)|^p\,dx\right\}^{1/p}\le Mt^\alpha,\qquad t>0 \tag{9.3}$$

with $A_t:=[a,b-t]$, if $A=[a,b]$, $t<b-a$, $A_t:=A$ in other cases. Here

$$|f|_{\operatorname{Lip}(\alpha,L_p)}:=\sup_{t>0}\big(t^{-\alpha}\omega(f,t)_p\big)\,. \tag{9.4}$$

We have restricted ourselves to values $0<\alpha\le1$, since for $\alpha>1$, $p\ge1$ the only functions for which (9.4) is finite are constants on A (see Proposition 7.1).

The importance of Lipschitz spaces related to the moduli of smoothness has been for the first time recognized by Zygmund [1945]. There are two ways to define such spaces. The first method is to require that a derivative $f^{(r)}$ of f should belong to a Lipschitz space. For $\alpha>0$ we write $\alpha=r+\beta$, where $r=0,1,\ldots$ is an integer and $0<\beta\le1$. For $0<p\le\infty$ we put the $\operatorname{Lip}(\alpha,L_p):=W^r\operatorname{Lip}(\beta,L_p)$. This space consists of all functions f for which $f,\ldots,f^{(r-1)}$ are locally absolutely continuous on A and satisfy $f^{(r)}\in\operatorname{Lip}(\beta,L_p)$. The seminorm is then

$$|f|_{\operatorname{Lip}(\alpha,L_p)}:=|f^{(r)}|_{\operatorname{Lip}(\beta,L_p)}=\sup_{t>0}\big(t^{-\beta}\omega(f^{(r)},t)_p\big)\,. \tag{9.5}$$

The other way is to replace Δ_tf in (9.3) by a difference of higher order. To define a *generalized Lipschitz space* $\operatorname{Lip}^*(\alpha,p)$, $\alpha>0$, $0<p\le\infty$, we denote

by r the smallest integer $r > \alpha$, that is, $r = [\alpha] + 1$ and require for $f \in L_p(A)$ that

$$(9.6) \qquad \|\Delta_t^r(f,\cdot)\|_p = \left\{\int_{A_{rt}} |\Delta_t^r(f,x)|^p\,dx\right\}^{1/p} \le Mt^\alpha\,, \qquad t > 0\,,$$

for some constant $M < \infty$. The seminorm is then

$$|f|_{\mathrm{Lip}^*(\alpha,L_p)} := \sup_{t>0}\big(t^{-\alpha}\omega_r(f,t)_p\big)\,.$$

By (7.13) the space $\mathrm{Lip}^*(\alpha, L_p)$ contains $\mathrm{Lip}(\alpha, L_p)$. For $0 < \alpha < 1$, the spaces coincide, (for $p = \infty$, it is necessary to replace L_∞ by the space $\widetilde{C}$ of uniformly continuous function on A). We see then that both spaces are identical with $\mathrm{Lip}\,\alpha$ of (9.1).

For $\alpha = 1$, $p = \infty$, we have $\mathrm{Lip}(1, \widetilde{C}) = \mathrm{Lip}\,1$, but $\mathrm{Lip}^*(1,\widetilde{C}) =: Z$ is the Zygmund space, which is characterized by (9.6) with $r = 2$, that is, by $|f(x+2t) - 2f(x+t) + f(x)| \le Mt$.

Example 1. The function $g(x) := x\log x$, $0 < x \le 1$, $:= 0$ for $x = 0$ belongs to $Z[0,1]$, but not to $\mathrm{Lip}\,1$. Indeed, $g''(x) = 1/x > 0$, $x > 0$, so that

$$0 \le \Delta_t^2(g,x) = 2t\log\frac{x+2t}{x+t} - x\log\left(1+\frac{t}{x}\right) + x\log\left(1+\frac{t}{x+t}\right)$$
$$\le 2t\log\frac{x+2t}{x+t} \le 2(\log 2)t\,.$$

Example 2. The singular point (point 0 above) can be inside $[a, b]$. Indeed, *if $g \in Z[0,1]$ and if g is odd on $[-1,1]$, then $g \in Z[-1,1]$*. The only interesting case is when not all three points x, $x+t$, $x+2t$ are in the same subinterval $[-1,0]$ or $[0,1]$. Let x be in the first, $x+t$, $x+2t$ in the second interval. Then $0 \le -x \le t$ and we have

$$\Delta_t^2(g,x) = \big[g(x+2t) - 2g(t) + g(-x)\big] - 2\big[g(-x) - 2g(\tfrac{1}{2}t) + g(x+t)\big]$$
$$+ 2\big[g(t) - 2g(\tfrac{1}{2}t) + g(0)\big]$$
$$= O(x+t) + O(x+\tfrac{t}{2}) + O(\tfrac{t}{2}) = O(t)\,.$$

We shall discuss relations between different spaces. In particular, we shall describe $\mathrm{Lip}(\alpha, L_p)$ in simpler terms. The following results hold for all of our standard sets A.

Theorem 9.1. *If $\alpha > 0$ is not an integer and $1 \le p \le \infty$, the relation $f \in \mathrm{Lip}(\alpha, L_p)$ holds if and only if f is a.e. equal to a function in $\mathrm{Lip}^*(\alpha, p)$, with equivalence of their semi-norms.*

This shall be proved as Corollary 3.2 in Chapter 6.

Besides the facts from §7, we need the lemma below, and the seminorm $\mathrm{Var}^* f$ for functions $f \in BV(A)$ (see §1). The latter is the variation of any *correction* $\bar f$ of f, which is equal to f except for a countable set of points, whose value $\bar f(c)$ lies between $\bar f(c+)$ and $\bar f(c-)$ at each interior point c of A, and which is continuous at the endpoints of A.

Lemma 9.2. *Each function $f \in BV$ belongs to* $\mathrm{Lip}(1, L_1)$, *and*

$$(9.7) \qquad |f|_{\mathrm{Lip}(1,L_1)} \le \mathrm{Var}^* f\,.$$

Proof. We treat only the case $A = [a, b]$, leaving the other cases to the reader. First of all, if $f \in W_1^1(A)$, then by changing the order of integration in the integral below, we get

$$\int_a^{b-h} |\Delta_h(f,x)|\,dx \le \int_a^{b-h}\int_x^{x+h} |f'(t)|\,dt\,dx = \int_a^b\int_a^b |f'(t)|\chi_{[x,x+h]}(t)\,dt\,dx$$
$$= \int_a^b |f'(t)| \int_a^b \chi_{[x,x+h]}(t)\,dx\,dt \le h\int_a^b |f'(t)|\,dt$$
$$= h\,\mathrm{Var}^* f\,,$$

which leads to (9.7). The general case follows from this and the fact that $f \in BV$ can be approximated by functions $g \in W_1^1$, so that $\int_a^b |f-g|\,dx < \varepsilon$ and $\mathrm{Var}\,g \le \mathrm{Var}^* f$. Indeed, we can assume that f is corrected and take the Steklov integral g_h of (4.18). We have $g_h \to f$ for $h \to 0+$ in L_1. Proving the second property of the g_h, we may assume that f is increasing (otherwise, we consider its positive and negative variation functions). Since $g_h'(x) = h^{-1}[f(x+h) - f(x)]$ a.e. on $[a, b-h]$, $= 0$ on $[b-h, b]$, we have

$$\mathrm{Var}\,g_h = \int_a^{b-h} g_h'(x)\,dx$$
$$= h^{-1}\left\{\int_{b-h}^b f\,dx - \int_a^{a+h} f\,dx\right\} \le f(b) - f(a) = \mathrm{Var}^* f\,. \qquad \square$$

Theorem 9.3. *For $r = 1, 2, \ldots,$ and $A = \mathbb{R}$, $\mathbb{R}_+$, $\mathbb{T}$, or $[a, b]$, the relation $f \in \mathrm{Lip}(r, L_p)$ holds if and only if f can be corrected on a set of measure zero to be a function g in W_p^r for $1 < p \le \infty$, or in $W^{r-1}(BV)$ for $p = 1$. Moreover, the seminorm $|f|_{\mathrm{Lip}(r,L_p)}$ is equal to the smallest M in the inequality*

$$(9.8) \qquad \omega_r(f,t)_p \le Mt^r\,, \qquad t > 0$$

and, respectively, to $\|g^{(r)}\|_p$ for $1 < p \le \infty$ or to $\mathrm{Var}^*\, g^{(r-1)}$ for $p = 1$.

With the obvious identification of functions f of an equivalence class with a single function g we have

$$(9.9) \quad \mathrm{Lip}(\alpha, L_p) = \mathrm{Lip}^*(\alpha, L_p) \ \text{ if } \ 1 \le p \le \infty \ \text{ and } \ \alpha > 1 \ \text{ is not an integer,}$$

(9.10)
$$\mathrm{Lip}(r, L_p) = W_p^r \ \text{ if } \ 1 < p \le \infty\,, \ = W^{r-1}(BV) \ \text{ if } \ p = 1\,, \ r = 1, 2, \ldots\,.$$

Proof of Theorem 9.3. First let $1 < p \le \infty$, and $f \in \mathrm{Lip}(r, L_p)$. If M is the smallest constant in (9.8), then

$$M \le |f|_{\mathrm{Lip}(r,L_p)} \le |f|_{W_p^r} . \tag{9.11}$$

The first inequality follows from $\omega_r(f,t)_p \le t^{r-1}\omega(f^{(r-1)},t)_p$ of (7.13), and the second from $\omega(f^{(r-1)},t)_p \le t|f|_{W_p^r}$ of (7.12).

It remains to prove that $\|f^{(r)}\|_p \le M$, if f satisfies (9.8). In this case, the functions $g_h(x) := h^{-r}\Delta_h^r f(x)$, $x \in A(rh)$, $:= 0$ otherwise, satisfy $\|g_h\|_p \le M$. Since $L_p(A)$ is a conjugate space, each bounded set in this space has a weakly $*$-convergent sequence. Thus for some $h_n \to 0$, the functions g_{h_n} converge weak $*$ in L_p to some g, $\|g\|_p \le M$. In particular, for all $\varphi \in C_0^\infty(A)$,

$$\begin{aligned}\int_A \varphi g\, dx &= \lim_{n\to\infty}\int_A \varphi g_{h_n}\, dx = \lim_{n\to\infty}\int_A h_n^{-r}\Delta_{-h_n}^r(\varphi) f\, dx \\ &= (-1)^r \int_A \varphi^{(r)} f\, dx .\end{aligned} \tag{9.12}$$

From Theorem 5.4 we see that g is a generalized derivative of f, $g = f^{(r)}$ a.e., and hence

$$|f|_{W_p^r} = \|g\|_p \le M .$$

Let now $p = 1$. If $f \in W^{r-1}(BV)$, then in analogy to (9.11),

$$M \le |f|_{\mathrm{Lip}(r,L_1)} \le |f^{(r-1)}|_{\mathrm{Lip}(1,L_1)} \le \mathrm{Var}^* f^{(r-1)} ; \tag{9.13}$$

the last inequality follows from (9.7). To invert this, we use the space $BV(A)$, and not $L_1(A)$, which is not a conjugate space. For the functions g_h, by (9.8), $\|g_h\|_1 \le M$. The functions $G_h(x) = \int_c^x g_h(t)\, dt$, where $c \in A$ is fixed, belong to BV with $\mathrm{Var}\, G_h \le M$. If $G_{h_n} \to G$ weak $*$ in BV, $\mathrm{Var}^* G \le M$, then, as in (9.12), $\int_A \varphi\, dG = (-1)^r \int_A \varphi^{(r)} f\, dx$ for all $\varphi \in C_0^\infty(A)$. Since $\int_A \varphi\, dG = -\int_A g\varphi'\, dx$, G is a generalized $r-1$-st derivative of f, $f^{(r-1)} = G$ a.e. Hence f can be corrected on a set of measure zero to be in $W^{r-1}(BV)$ and to satisfy $\mathrm{Var}^* f^{(r-1)} \le M$. □

§ 10. Besov Spaces

The characterization of functions f which have a given upper bound for the error of approximation sometimes requires a finer scale of smoothness than is provided by the Lipschitz spaces. For each $\alpha > 0$, this can be accomplished by introducing a second parameter q and applying α, q norms (rather than α, ∞ norms) to the modulus of smoothness $\omega_r(f,\cdot)$ of f. Let $\alpha > 0$ be given and let $r := [\alpha] + 1$. For $0 < p, q \le \infty$, the *Besov space* $B_q^\alpha(L_p)$ is the collection of all functions $f \in L_p(A)$ such that

(10.1)

$$|f|_{B_q^\alpha(L_p)} := \|\omega_r(f,\cdot)\|_{\alpha,q} = \begin{cases} \left(\displaystyle\int_0^\infty [t^{-\alpha}\omega_r(f,t)_p]^q \frac{dt}{t}\right)^{1/q}, & 0 < q < \infty, \\ \sup_{t>0} t^{-\alpha}\omega_r(f,t)_p , & q = \infty \end{cases}$$

is finite. They have been introduced by Besov. The quantity (10.1) is a seminorm if $1 \le p, q \le \infty$ (a quasi-seminorm in the other cases). The (quasi) norm for $B_q^\alpha(L_p)$ is

$$\|f\|_{B_q^\alpha(L_p)} := \|f\|_p + |f|_{B_q^\alpha(L_p)} . \tag{10.2}$$

In particular, for $q = \infty$, $B_\infty^\alpha(L_p) = \mathrm{Lip}^*(\alpha, L_p)$.

We shall frequently use the following facts concerning the definition of Besov spaces.

Theorem 10.1. *Let $\alpha > 0$, $0 < q, p \le \infty$. (i) For compact A, a seminorm for $B_q^\alpha(L_p)$, equivalent to the one defined by* (10.1), *is obtained by replacing the integral (or the supremum) over $(0,\infty)$ by one over $(0,1)$. For all A, this change in* (10.1) *gives a (quasi) norm on $B_q^\alpha(L_p)$ equivalent to* (10.2).

(ii) *For all A, one obtains a (quasi) norm for $B_q^\alpha(L_p)$ equivalent to* (10.1) *by replacing ω_r by ω_k, $k > r$.*

Proof. We treat only the case $1 \le p \le \infty$; similar arguments apply when $0 < p < 1$. (i) Let $\varphi(t) := \omega_r(f,t)_p$. If A is finite then φ is constant for $t \ge |A|$ and therefore $\varphi(t) \le \varphi(|A|)$. Because of (7.8) (and the remark following (7.7) in the case $p < 1$), this implies $\varphi(\frac12) \le \varphi(t) \le \varphi(|A|) \le C\varphi(\frac12)$, for $\frac12 \le t < \infty$. Therefore, if $q < \infty$,

$$\int_1^\infty [t^{-\alpha}\varphi(t)]^q \frac{dt}{t} \le C^q(\alpha q)^{-1}\varphi(\tfrac12)^q \le C_1^q \int_{1/2}^1 [t^{-\alpha}\varphi(t)]^q \frac{dt}{t} ,$$

and (i) follows. This argument applies also for $q = \infty$. For non-compact A, we can replace the integral (or supremum) over $[1,\infty)$ by $C\|f\|_p$ by invoking the inequality $\omega_r(f,t)_p \le 2^r\|f\|_p$.

(ii) We denote by $|\cdot|_r$ and $|\cdot|_k$ the seminorms (10.1) corresponding to ω_r and ω_k respectively. By (7.5), $\omega_k(f,t)_p \le 2^{k-r}\omega_r(f,t)_p$; hence $|\cdot|_k \le C|\cdot|_r$, with the same inequality for the (quasi) norms (10.2). For the inequality in the opposite direction, we employ Marchaud's Theorem 8.1. This yields for $\varphi(s) := s^{-r}\omega_k(f,s)_p$,

$$t^{-\alpha}\omega_r(f,t)_p \le C_1 t^{r-\alpha}\int_t^\infty \varphi(s)\frac{ds}{s} + C_2 t^{r-\alpha}\|f\|_p , \tag{10.3}$$

where $C_2 = 0$ if A is not a compact interval, and $0 < t \le 1$ if A is compact. To estimate $\|\omega_r(f,\cdot)_p\|_{\alpha,q}$, we apply the functional $\int_0^\infty [\,\cdot\,]^q \frac{dt}{t}$ to the right side of (10.3). If the term $\|f\|_p$ is present, we are allowed, by (i), to replace the interval of integration by $[0,1]$. The function $t^{(r-\alpha)q-1}$ is integrable and we obtain at most $C\|f\|_p$. The contribution of the first term on the right in (10.3), by Hardy's inequality (3.2) with $\theta := r - \alpha$ is at most $C\|\omega_k(f,\cdot)\|_{\alpha,q}$. □

It is sometimes useful to discretize the seminorms (i) and (ii) of the theorem. From property (7.8) and the monotonicity of ω_r, we have for $1 \le p \le \infty$, and $t \in [2^{-k-1}, 2^{-k}]$,

$$2^{-r}\varphi(2^{-k}) \le \varphi(t) \le 2^{\alpha}\varphi(2^{-k})$$

with $\varphi(t) := t^{-\alpha}\omega_r(f,t)_p$. It follows that

$$\left(\int_{2^{-k-1}}^{2^{-k}} \varphi(t)^q \frac{dt}{t}\right)^{1/q} \sim \varphi(2^{-k}), \qquad k \in \mathbb{Z} \tag{10.4}$$

with the constants of equivalency independent of f and k. A similar argument applies if $q = \infty$ and also in the case $0 < p < 1$. Using the seminorm (i), we obtain for finite A:

$$|f|_{B_q^\alpha(L_p)} \sim \begin{cases} \left(\sum_{k=0}^{\infty} [2^{k\alpha}\omega_r(f,2^{-k})_p]^q\right)^{1/q}, & 0 < q < \infty, \\ \sup_{k\ge 0}\{2^{k\alpha}\omega_r(f,2^{-k})_p\}, & q = \infty \end{cases} \tag{10.5}$$

with constants of equivalency independent of f.

This argument shows that when A is non-compact, we obtain an equivalent seminorm for $B_q^\alpha(L_p)$ by taking the sum (or supremum) in (10.5) over $k \in \mathbb{Z}$. Moreover, because of (i), for any A, we obtain an equivalent norm for $B_q^\alpha(L_p)$ by adding $\|f\|_p$ to the right side of (10.5). This argument applies verbatim to discretizing (ii).

For fixed α and p, the space $B_q^\alpha(L_p)$ gets larger with increasing q and we have the embedding inequalities

$$|f|_{B_{q_1}^\alpha(L_p)} \le C|f|_{B_q^\alpha(L_p)}, \qquad 0 < q < q_1 \le \infty. \tag{10.6}$$

Indeed, this follows from (10.5) (or the remark following (10.5) in the case $|A| = \infty$) because the l_q norms decrease with increasing q.

If $\beta < \alpha$, then for all $0 < q, q_1 \le \infty$, we have $B_q^\alpha(L_p) \subset B_{q_1}^\beta(L_p)$ and

$$\|f\|_{B_{q_1}^\beta(L_p)} \le C\|f\|_{B_q^\alpha(L_p)}. \tag{10.7}$$

For the proof, we first note that both norms can be computed with the same r because of (ii) of Theorem 10.1. For this value of r, we let $|\cdot|_1$ and $|\cdot|_2$ be the seminorms given by the right side of (10.5) for α, q and β, q_1, respectively. Because of (10.5) (and the remark following (10.5) in the case $|A| = \infty$), it is enough to show that

$$|\cdot|_2 \le C|\cdot|_1. \tag{10.8}$$

Now, $2^{k\beta} \le 2^{k\alpha}$, $k = 0, 1, \dots$. Hence we have (10.8) for $q \le q_1$ because the l_{q_1} norm does not exceed the l_q norm. If $q_1 < q$, we have

$$\begin{aligned}\left(\sum_{k=0}^{\infty}[2^{k\beta}\omega_r(f,2^{-k})_p]^{q_1}\right)^{1/q_1} &\le \left(\sum_{k=0}^{\infty} 2^{k(\beta-\alpha)q_1}\right)^{1/q_1} \sup_{k\ge 0}\{2^{k\alpha}\omega_r(f,2^{-k})_p\} \\ &\le C\sup_{k\ge 0}\{2^{k\alpha}\omega_r(f,2^{-k})_p\} \le C\|2^{k\alpha}\omega_r(f,2^{-k})_p\|_{l_q}\end{aligned}$$

because the l_∞ norm does not exceed the l_q norm.

Further properties of Besov spaces are given in Chapter 6.

§ 11. Problems and Notes

Problem 11.1. If f is a measurable function on $A = [a,b]$, then $\lim_{p\to\infty}\|f\|_p = \|f\|_\infty$, and $\lim_{p\to 0}\|f\|_p = \|f\|_0$, where

$$\|f\|_0 = \exp\int_A \log|f(x)|\,dx. \tag{11.1}$$

Problem 11.2. If $r = 1, 2, \dots$, and $f^{(r)} \in BV[a,b]$, then there are functions $f_n \in C^\infty$ with the properties $\|f - f_n\|_{W_1^r} \to 0$ and $\|f_n^{(r+1)}\|_1 \le \operatorname{Var} f^{(r)}$. If $f^{(r)}$ is continuous, then one can assume that $f_n^{(r)}$ converges to $f^{(r)}$ uniformly on $[a,b]$.

This can be used in the proof of Lemma 9.2.

Note 11.1. The spaces $M_\alpha[0,1]$, $\Lambda_\alpha[0,1]$ for $0 < \alpha < 1$ have interesting extremal properties. They are dual to each other in the sense that for a function f in one of the spaces, its norm $\|f\| = \sup_{\|g\|\le 1}\int_A fg\,dx$, with the norm of g in the other space. For a rearrangement invariant space X on A, the norm of the characteristic function χ_B, $B \subset A$ depends only on $u = |B|$; this is the fundamental function $\varphi_X(u)$ of X. All three spaces Λ_α, M_α and L_p, $p = 1/\alpha$ on A have the same fundamental function $\varphi(u) = u^\alpha$; M_α is the largest, Λ_α the smallest of all spaces with this $\varphi(u)$.

11.2. Theorem 4.4 has a converse. We call a Banach function space X, satisfying the continuous imbedding $L_1 \cap L_\infty \subset X \subset L_1 + L_\infty$, an "exact interpolation space" between L_1 and L_∞ if each linear operator U on $L_1 + L_\infty$, which has the properties

$$\|U\|_{L_1} \le M, \qquad \|U\|_{L_\infty} \le M \tag{11.2}$$

also maps X into itself with norm $\le M$. With this definition, Calderón [1966] proves: *X is an exact interpolation space between L_1 and L_∞ if and only if it is rearrangement-invariant.*

11.3. Properties (11.2) are closely related to the relation $Uf \prec Mf$, proved in (4.15): This relation holds for all $f \in L_1 + L_\infty$ if and only if (11.2) is valid.

Chapter 3. Best Approximation

§ 1. Introduction: Existence of Best Approximation

This chapter is somewhat outside of the main flow of the book: it deals with *best approximation*, while the rest, in the main, is devoted to *sufficiently good approximation.* The roots of Chapter 3 lie in the past. It was one of the first great Russian mathematicians, P. L. Chebyshev (see [B-1962]), who in the latter half of the 19th century studied polynomials $P \in P_n$ of best uniform approximation to functions $f \in C[a,b]$. The key theorem of the present chapter is

Theorem A. *If $f \notin \mathcal{P}_n$, then $P \in \mathcal{P}_n$ is the polynomial of best uniform approximation to f if and only if there are $n+2$ points x_j, $a \le x_1 < \cdots < x_{n+2} \le b$, so that the difference $f(x) - P(x)$ takes the values $\pm\|f - P\|$ with alternating signs for $x = x_j$, $j = 1, \ldots, n+2$.*

This theorem is traditionally called the *Chebyshev theorem,* although he did not prove it. For the proof, one had to wait until the first years of the 20th century when it was stated and proved independently by the American Blichfeldt [1901] and by Kirchberger [1902], a Ph.D. student of Hilbert. Chebyshev himself did not discuss problems of existence and characterization of best approximants P. He merely proved that if f is differentiable, then there are at least $n+2$ critical points (endpoints of the interval and points with $f'(x) - P'(x) = 0$). He also found P explicitly if $f(x) = x^{n+1}$ (Chebyshev polynomials of §6). The proof of Theorem A is beautiful and simple, but it allows many useful generalizations, which may hide its simplicity.

More than half of Chapter 3 is devoted to the study of *Haar spaces* $\Phi_n \subset C(A)$ on $A = [a,b]$ or $A = \mathbb{T}$. They are characterized by the property that each $f \in C(A)$ has exactly one element of best uniform approximation from Φ_n. The importance of weak Haar spaces (§12), introduced by Karlin and Studden [A-1966], lies in the fact that the Schoenberg spline spaces of Chapter 5 belong to this category.

The main results of the present chapter concern uniform approximation. There are also theorems about L_p-approximation. The best of them are for the L_1-approximation of continuous functions.

Let X be a Banach space with real or complex scalars and Y be a closed linear subspace of X. We recall that for each $f \in X$, the error of approximation

$E(f)$ of f by elements from Y is

$$E(f) := E(f,Y)_X := \inf_{P \in Y} \|f - P\| . \tag{1.1}$$

The error $E(f)$ is a continuous function of f. Indeed, for any $f, f_0 \in X$,

$$E(f_0) - \|f - f_0\| \le E(f) \le E(f_0) + \|f - f_0\| \tag{1.2}$$

and therefore $E(f) \to E(f_0)$, if $f \to f_0$.

If the infimum in (1.1) is attained for some $P = P_0$, this P_0 is called *a best approximation to f from Y*. We let $B(f) \subset Y$ be the set of all elements $P \in Y$ of best approximation to f. Clearly, $B(f)$ is *closed.* It is also *convex.* Indeed, if $P_1, P_2 \in B(f)$, $\alpha_1, \alpha_2 \ge 0$, $\alpha_1 + \alpha_2 = 1$, then

$$E(f) \le \|f - (\alpha_1 P_1 + \alpha_2 P_2)\| \le \alpha_1 \|f - P_1\| + \alpha_2 \|f - P_2\| = E(f)$$

so that $\alpha_1 P_1 + \alpha_2 P_2 \in B(f)$.

In this chapter we shall discuss questions of existence and uniqueness of best approximations and discuss their characteristic properties. We shall use the notation X_n for a finite dimensional subspace of X (of dimension n) and write $E_n(f)$ for $E(f)$ when we wish to emphasize the dependence on n.

Theorem 1.1. *For each finite dimensional subspace X_n of X and each $f \in X$, there is a best approximation to f from X_n.*

Proof. There is a sequence $P_m \in X_n$ for which $\|P_m - f\| \to E(f)$. Since $\|P_m\| \le \|f\| + \|f - P_m\|$, (P_m) is a bounded sequence in the finite dimensional space X_n, hence it is precompact. Therefore, there is a subsequence (P_{m_j}) and a $P_0 \in X_n$ with $\|P_{m_j} - P_0\| \to 0$, $j \to \infty$. Then $\|f - P_{m_j}\| \to \|f - P_0\|$ and therefore $\|f - P_0\| = E(f)$. □

For some Banach spaces X, we can claim that a best approximation is unique for each closed linear subspace $Y \subset X$: each $f \in X$ has at most one element of best approximation in Y. This is the case for *strictly convex spaces,* which are characterized by the following property:

$$\begin{cases} f_1 \ne f_2\,, & \|f_1\| = \|f_2\| = 1\,, \quad \alpha_1, \alpha_2 > 0\,, \quad \alpha_1 + \alpha_2 = 1 \\ \text{imply} & \|\alpha_1 f_1 + \alpha_2 f_2\| < 1\,. \end{cases} \tag{1.3}$$

If X is strictly convex and $P_1, P_2 \in Y$ both satisfy

$$\|f - P_1\| = \|f - P_2\| = E(f) > 0\,,$$

then, using (1.3), we obtain a contradiction:

$$E(f) \le \left\| f - \frac{1}{2}(P_1 + P_2) \right\| = \left\| \frac{1}{2}(f - P_1) + \frac{1}{2}(f - P_2) \right\| < E(f)\,,$$

unless $P_1 = P_2$.

It is easy to check that the spaces L_p, $1 < p < \infty$, are strictly convex. For this purpose, it is sufficient to know the cases of equality in the inequality of Minkowski. However, the important spaces L_1 and C do not have this property, and the uniqueness statement does not apply to them. For example, take the space $C[0,1]$. Let $f(x) = 1$, and $f_1(x) = x$, $0 \le x \le 1$. We wish to approximate f by functions $P = a_1 f_1$. For each P, $\|f - P\| \ge f(0) = 1$. On the other hand, this lower bound for $\|f - P\|$ is attained by $P = a_1 f_1$, $0 \le a_1 \le 2$. Hence, all these P are elements of best approximation. Likewise, for the space $L_1[0,2]$ with Lebesgue measure, let $f(x) = 1$ on $[0,1], = 0$ on $(1,2]$, and $f_1(x) \equiv 1$. Then, $\|f - a_1 f_1\|_1 = |1 - a_1| + |a_1|$ attains its minimum $= 1$ for all $0 \le a_1 \le 1$.

In the space $L_1(A, d\mu)$ of μ-integrable functions on A the following remark is often useful: *If P_1, P_2 are two best approximations to f, then $f(x) - P_1(x)$ and $f(x) - P_2(x)$ can be of opposite sign only on a set of measure zero.* Since $\frac{1}{2}(P_1 + P_2)$ is also a best approximation, this follows from

(1.4)
$$\int_A \{|f(x) - P_1(x)| + |f(x) - P_2(x)| - |2f(x) - P_1(x) - P_2(x)|\}\, d\mu(x) = 0 .$$

Here is another useful remark. Let X be some Banach space of functions on an interval $[-a,a]$ or on $\mathbb{T}$ for which $\|f(-\cdot)\| = \|f(\cdot)\|$. We want to approximate $f \in X$ by algebraic or trigonometric polynomials of degree $\le n$. Then an even (odd) function has an even (odd) best approximation. For example, an even function $f(x)$ together with $P_n(x)$, has also $P_n(-x)$ and therefore $\frac{1}{2}(P_n(x) + P_n(-x))$ as its polynomials of best approximation.

If best approximation exists and is unique, then it is continuous. Let X be a Banach space and let $Y \subset X$ be a *set of unicity*. This means that each $f \in X$ has a unique best approximation $P(f) := P \in Y$.

Theorem 1.2. *The operator P of best approximation from a finite dimensional unicity subspace X_n is continuous: $f \to f_0$ in X implies $Pf \to Pf_0$.*

Proof. Let $f_k \to f$ in the space X. We prove that $P(f_k) \to P(f)$. The elements f_k are uniformly bounded in X, $\|f_k\| \le M$ and since

$$\|P(f_k)\| \le \|P(f_k) - f_k\| + \|f_k\| \le 2\|f_k\| \le 2M ,$$

the norms $\|P(f_k)\|$ are also bounded. Assume that $P(f_k) \not\to P(f)$. Since the space X_n is locally compact, there exists a subsequence $P(f_{k'}) \to P_0 \ne Pf$. Now $P(f_{k'})$ approximates $f_{k'}$ at least as well as Pf does:

$$\|P(f_{k'}) - f_{k'}\| \le \|P(f) - f_{k'}\| .$$

Letting $k' \to \infty$, we obtain that P_0 is also a best approximation to f, $P_0 = P(f)$ which is the desired contradiction. □

There is one case where it is particularly simple to determine best approximants. Let H be a Hilbert space with inner product $(\cdot,\cdot)$ and let H_0 be one

of its subspaces. An element $g \in H_0$ is a best approximation to $f \in H$ if and only if the following orthogonality conditions hold:

$$(f - g, h) = 0, \quad \text{for all } h \in H_0 . \tag{1.5}$$

Indeed, for any $g, h \in H_0$ and any real number $\alpha \ne 0$,

$$\|f - g - \alpha h\|^2 = \|f - g\|^2 - 2\alpha(f - g, h) + \alpha^2\|h\|^2 . \tag{1.6}$$

If g is a best approximation to f and (1.5) fails to hold for some h, then we take α in (1.6) small and of the same sign as $(f - g, h)$ and obtain the contradiction $\|f - g - \alpha h\| < \|f - g\|$. On the other hand if (1.5) is satisfied then (1.6) gives $\|f - g - h\| > \|f - g\|$, for all $h \in H_0$, and g is therefore a best approximation. This also shows that g is unique whenever it exists.

Orthogonality is also useful in a general Banach space X. We will consider the *bounded linear* functionals on X. Such a functional λ is *orthogonal* to the subspace Y if $\lambda(P) = 0$ for all $P \in Y$. We denote by $Y^\perp$ the set of all such λ. The following remark is trivial but often useful:

(1.7). *Let Y be a closed linear subspace of X, and $f \in X$. If $\lambda \in Y^\perp$ has norm one, then $E(f) \ge \lambda(f)$.*

Indeed, then for all $g \in Y$, $\|f - g\| \ge \lambda(f - g) = \lambda(f)$.

Moreover, the Hahn-Banach theorem yields the existence of a functional λ for which

$$\|\lambda\| = 1, \quad \lambda \in Y^\perp, \text{ and } \lambda(f) = E(f) . \tag{1.8}$$

From these two statements we derive:

Theorem 1.3. *Let Y be a closed subspace of X, and $f \in X \backslash Y$. Then $g_0 \in Y$ is a best approximation to f from Y if and only if there is a bounded linear functional $\lambda \in Y^\perp$ with the properties $\|\lambda\| = 1$ and $\|f - g_0\| = \lambda(f)$. Moreover,*

$$E(f) = \sup_{\substack{\lambda \in Y^\perp \\ \|\lambda\|=1}} \lambda(f) . \tag{1.9}$$

There is a similar result to Theorem 1.3 in which Y is replaced by a general convex set G. Its proof depends on the separation of convex sets by means of hyperplanes (that is, by means of sets $\{x : \lambda(x) = a\}$) (see the book of Royden [B-1968, p. 204]).

Theorem 1.4. *If G and G_0 are two disjoint convex sets in a normed linear space X and if G_0 is open, then there exists a linear functional λ on X and $\gamma \in \mathbb{R}$ for which*

$$\lambda(g) < \gamma \le \lambda(f), \quad f \in G,\ g \in G_0 . \tag{1.10}$$

We shall now derive a formula similar to (1.9) when the *linear subspace* of X is replaced by a *convex set* $G \subset X$. Let

$$E(f) := E_G(f) := \inf_{g \in G} \|f - g\|, \quad f \in X.$$

Theorem 1.5 (Fenchel [1949]). *If G is a convex subset of a normed linear space X, then*

$$E(f) = \sup_{\substack{\lambda \in X^* \\ \|\lambda\| \le 1}} \left\{ \lambda(f) - \sup_{g \in G} \lambda(g) \right\}, \quad f \in X, \tag{1.11}$$

where λ are bounded linear functionals of norm ≤ 1 on X. Moreover, the supremum is attained for some $\lambda_0 \in X^*$, $\|\lambda_0\| = 1$.

Proof. Let $N(f)$ be the right-hand side of (1.11). For arbitrary $g' \in G$ and $\lambda \in X^*$, $\|\lambda\| \le 1$,

$$\|f - g'\| \ge \lambda(f) - \lambda(g') \ge \lambda(f) - \sup_g \lambda(g).$$

Taking here an infimum over all g', a supremum over all λ, we obtain $E(f) \ge N(f)$.

If $f \in \overline{G}$, then $E_G(f) = 0$, but by what we have already proved also $N(f) = 0$. Therefore, without loss of generality, let $f \in X \setminus \overline{G}$. Then $E(f) =: r > 0$. The set G and the ball $B_r(f)$ in X with center f and radius r are disjoint convex sets, and the second set is the closure of its interior. From Theorem 1.4, there exists a non-zero linear functional $\lambda_0 \in X^*$, which we can assume to have norm $\|\lambda_0\| = 1$ so that $\lambda_0(g) \le \lambda_0(f')$ for all $g \in G$, $f' \in B_r(f)$. Then

$$\sup_g \lambda_0(g) \le \inf_{\|f'-f\| \le r} \lambda_0(f') = \inf_{\|f'\| \le r} \lambda_0(f - f') = \lambda_0(f) - \sup_{\|f'\| \le r} \lambda_0(f').$$

Consequently

$$E(f) = r\|\lambda_0\| = \sup_{\|f'\| \le r} \lambda_0(f') \le \lambda_0(f) - \sup_{g \in G} \lambda_0(g) \le N(f), \tag{1.12}$$

so that $E(f) = N(f)$. The inequality (1.12) yields now that λ_0 is an extreme functional which realizes the supremum in (1.11). □

One may also obtain Theorem 1.3 as an immediate corollary of Theorem 1.5. For if $\lambda \notin Y^\perp$, then for some $g_0 \in Y$, $\lambda(g_0) \ne 0$, and then $\sup_{g \in Y} \lambda(g) = \sup_{c \in \mathbb{R}} \lambda(c g_0) = \infty$. Therefore, this λ can be omitted in the supremum over λ in (1.11).

§ 2. Kolmogorov's Theorem

In this and the next sections, we discuss approximation in the space $C(A)$ of continuous real or complex valued functions on the compact Hausdorff topological space A with the norm $\|f\| := \max_{x \in A} |f(x)|$. Let X_n be a given (complex or real) n-dimensional subspace of $C(A)$. We shall treat the "complex case" and the "real case" together.

A characterization of best approximation in $C(A)$ is given by the following theorem.

Theorem 2.1 (Kolmogorov [1948]). *A function P in X_n is a best approximation to $f \in C(A)$ if and only if for each $Q \in X_n$*

$$\max_{x \in A_0} \operatorname{Re} \left\{ [f(x) - P(x)] \overline{Q(x)} \right\} \ge 0, \tag{2.1}$$

where A_0 denotes the set (which depends on f and P) of all points $x \in A$ for which $|f(x) - P(x)| = \|f - P\|$.

Proof. Assume first that P is an element of best approximation. Let $\|f - P\| = E$. If (2.1) is not true, there exists a Q such that

$$\max_{x \in A_0} \operatorname{Re} \left\{ [f(x) - P(x)] \overline{Q(x)} \right\} = -2\varepsilon$$

for some $\varepsilon > 0$. By continuity, there exists an open subset G of A, $G \supset A_0$, such that

$$\operatorname{Re} \left\{ [f(x) - P(x)] \overline{Q(x)} \right\} < -\varepsilon, \quad x \in G.$$

Let us see how f is approximated by $P_1 := P - \lambda Q$, where $\lambda > 0$ is small. Let $M := \|Q\|$. First we assume that $x \in G$. In this case,

$$\begin{aligned} |f(x) - P_1(x)|^2 &= |[f(x) - P(x)] + \lambda Q(x)|^2 \\ &= |f(x) - P(x)|^2 + 2\lambda \operatorname{Re} \left\{ [f(x) - P(x)] \overline{Q(x)} \right\} + \lambda^2 |Q(x)|^2 \\ &< E^2 - 2\lambda\varepsilon + \lambda^2 M^2. \end{aligned}$$

If we take $\lambda < M^{-2}\varepsilon$, then $\lambda^2 M^2 < \lambda\varepsilon$, and we obtain

$$|f(x) - P_1(x)|^2 < E^2 - \lambda\varepsilon, \quad x \in G. \tag{2.2}$$

To estimate $f(x) - P_1(x)$ for $x \notin G$, we note that the complement of G is a closed set $F \subset A$, and that on F, $|f(x) - P(x)| < E$. Hence, for some $\delta > 0$, $|f(x) - P(x)| < E - \delta$, $x \in F$. If we take λ so small that $\lambda < (2M)^{-1}\delta$, we shall have

$$\begin{aligned} |f(x) - P_1(x)| &\le |f(x) - P(x)| + \lambda |Q(x)| \\ &\le E - \delta + \frac{1}{2}\delta = E - \frac{1}{2}\delta, \quad x \in F. \end{aligned} \tag{2.3}$$

From (2.2) and (2.3) we see that for all sufficiently small positive values of λ, P_1 approximates f better than P. Hence, the condition (2.1) is necessary.

To show that (2.1) is also sufficient, assume that the condition holds for each Q. Taking an arbitrary $P_1 \in X_n$, we see that there is a point $x_0 \in A_0$ such that for $Q := P - P_1$,

$$\operatorname{Re}\left\{[f(x_0) - P(x_0)]\overline{Q(x_0)}\right\} \geq 0 .$$

Then,

$$\begin{aligned} |f(x_0) - P_1(x_0)|^2 &= |f(x_0) - P(x_0)|^2 \\ &\quad + 2\operatorname{Re}\left\{[f(x_0) - P(x_0)]\overline{Q(x_0)}\right\} + |Q(x_0)|^2 \\ &\geq |f(x_0) - P(x_0)|^2 = \|f - P\|^2 , \end{aligned}$$

since $x_0 \in A_0$. Hence, P_1 cannot approximate f with an error less than $\|f-P\|$, and P must be a best approximation from X_n. □

This proof also works for the "real case". Condition (2.1) becomes (2.4) below, and we obtain

Theorem 2.2. *For real $C(A)$, the function $P \in X_n$ is a best approximation to f if and only if for each $Q \in X_n$,*

(2.4) $$\max_{x \in A_0}\{[f(x) - P(x)]Q(x)\} \geq 0 .$$

There is a useful interpretation of Kolmogorov's theorems in terms of orthogonal linear functionals. This uses properties of convex sets. If B is a subset of a Banach space X, its convex hull, B_c, consists of all $x \in X$ of the form

(2.5) $$x = \sum_{i=0}^{r} p_i x^{(i)}, x^{(i)} \in B, p_i \geq 0 , \quad \sum_{i=0}^{r} p_i = 1 , \quad r = 1, 2, \ldots .$$

For sets B in $\mathbb{R}^n$, we have a stronger statement:

Theorem 2.3 (Carathéodory). *The convex hull of a set $B \subset \mathbb{R}^n$ consists of all $x \in \mathbb{R}^n$ of the form (2.5) with $r \leq n$.*

Proof. Let $x \in B_c$ be given, and let r be the smallest integer for which a representation (2.5) is possible. We have to prove that $r \leq n$. Assume that $r > n$. Then the vectors $x^{(1)} - x^{(0)}, \ldots, x^{(r)} - x^{(0)}$ are linearly dependent, and there is a relation

$$\sum_{i=1}^{r} \beta_i \left[x^{(i)} - x^{(0)}\right] = \sum_{i=0}^{r} \alpha_i x^{(i)} = 0 , \quad \sum_{i=0}^{r} \alpha_i = 0 ,$$

where not all β_i and not all α_i are zero. Thus, in addition to (2.5), we have

$$x = \sum_{i=0}^{r} (p_i + \lambda\alpha_i)\, x^{(i)} .$$

We want to choose λ so that $q_i := p_i + \lambda\alpha_i \geq 0$ for all i and so that $q_i = 0$ for at least one i. There are α_i that satisfy $\alpha_i < 0$. It is easy to see that $\lambda := \min\{p_i/|\alpha_i| : \alpha_i < 0\} > 0$ has the required properties. Since $\Sigma q_i = 1$, we have obtained a representation of type (2.5) with at most r terms. This is a contradiction. □

This theorem holds also for subsets of the n-dimensional *complex* space $\mathbb{C}^n$, of points $z = (z_1, \ldots, z_n)$ with complex coordinates $z_k = x_k + iy_k$, which is isomorphic to the $2n$-dimensional, real space $\mathbb{R}^{2n}$. In this case, the restriction upon r in Theorem 2.3 becomes $r \leq 2n$.

A (real) linear functional $\lambda(x)$, $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$ is given by $\lambda(x) = \sum_1^n a_k x_k$. For $z \in \mathbb{C}^n$ this is replaced by the formula $\lambda(z) = \sum_1^n (a_k x_k + b_k y_k)$. We can combine the real and the complex case by writing

(2.6) $$\lambda(z) = \operatorname{Re}\left(\sum_1^n \bar{c}_k z_k\right) ,$$

where c_k are arbitrary complex (or real) numbers.

We return to a discussion of Theorem 2.1. Let $\phi_1, \ldots, \phi_n$ be a basis for X_n and let B be the subset of $\mathbb{C}^n$ (or $\mathbb{R}^n$) consisting of all points

$$z = \left\{(f(x) - P(x))\overline{\phi_1(x)}, \ldots, (f(x) - P(x))\overline{\phi_n(x)}\right\} , \quad x \in A_0 .$$

Because all these functions are continuous and the set A_0 is compact, B is also compact; hence its convex hull B_c is convex and compact in $\mathbb{C}^n$ (or $\mathbb{R}^n$).

A linear functional (2.6) on B has the form

$$\lambda(z) = \operatorname{Re}(f - P)(x)\overline{Q(x)}, \quad x \in A_0 ; \quad Q = \sum_1^n c_k \phi_k .$$

Therefore, condition (2.1) of Theorem 2.1 can now be written in the form $\max_{z \in B} \lambda(z) \geq 0$ for each λ. We claim that this condition is equivalent to $0 \in B_c$. Indeed, by Theorem 1.4, $0 \notin B_c$ means that there is a functional λ for which $\lambda(z) < 0$ for all $z \in B_c$, or equivalently, for all $z \in B$, that is, a functional with $\max_{z \in B} \lambda(z) < 0$.

On the other hand, by Carathéodory's theorem, $0 \in B_c$ means that $0 = \sum_{i=0}^r p_i z^{(i)}$, $z^{(i)} \in B$. Let x_i be points of A_0 which produce the $z^{(i)}$, $i = 0, \ldots, r$. The k-th coordinate of the last sum, $k = 1, \ldots, n$, is $\sum_{i=0}^r p_i [f(x_i) - P(x_i)]\overline{\phi_k(x_i)}$, and this must be zero. Equivalently,

(2.7) $$\sum_{i=0}^{r} p_i [f(x_i) - P(x_i)]\overline{Q(x_i)} = 0 , \text{ for each polynomial } Q.$$

We obtain:

Theorem 2.4 (Rivlin-Shapiro [1961]). *A function $P \in X_n$ is a best approximation to $f \in C(A)$ if and only if there exist points $x_i \in A_0$, $i = 0, \dots, r$, and numbers $p_i \geq 0$, $\sum_0^r p_i = 1$ so that (2.7) holds for all $Q \in X_n$. Here $r \leq n$ for the real case and $r \leq 2n$ for the complex case.*

Corollary 2.5. *If P is a best approximation to f on A, then P is also a best approximation to f on a certain finite subset of A_0, which consists of $r+1$ points ($r \leq n$, in the real case and $r \leq 2n$, in the complex case).*

From Theorem 2.4, we can derive that in Theorem 1.3, if $X = C(A)$ and $Y = X_n$ is finite dimensional, one needs to take into account only a few very special linear functionals λ^* which we call the approximation functionals of f and P. They are

$$(2.8) \quad \begin{cases} \lambda^*(g) := \sum_{i=0}^{r} \alpha_i \overline{g}(x_i), \\ \alpha_i := p_i \operatorname{sign} \overline{[f(x_i) - P(x_i)]}, \quad p_i \geq 0, \quad i = 0, \dots, r, \quad \sum_{i=0}^{r} p_i = 1, \end{cases}$$

where $x_0, \dots, x_r$ are in A_0 and $r \leq n$ ($r \leq 2n$ in the complex case). Indeed, if P is a best approximation to f, we have $\|\lambda^*\| = \sum_{i=0}^r |\alpha_i| = 1$ and $\lambda^*(f-P) = \|f-P\|$. Since $f(x_i) - P(x_i) = \|f-P\| \operatorname{sign}[f(x_i) - P(x_i)]$, (2.7) is equivalent to $\lambda^*(Q) = 0$. This leads to the following

Corollary 2.6. *Let $f \notin X_n$. Then P is an element of best approximation to f if and only if there exists an approximation functional (2.8) for this P and some points $x_0, \dots, x_r \in A_0$, $r \leq n$ which is orthogonal to X_n.*

It is instructive to compare Theorem 2.1 with the later Theorem 5.1 of Chebyshev. The latter is simpler and more concrete but only applies to special spaces X_n. The advantage of Kolmogorov's theorem lies in its adaptability to many different situations. Here is an example of *complex* approximation.

We consider the function $f(z) = (z - \alpha)^{-1}$ on the disk $A : |z| \leq 1$; the number α, $|\alpha| > 1$ is a complex constant. Let

$$(2.9) \qquad P_n(z) := \frac{1}{z-\alpha} - Cz^n \frac{\overline{\alpha}z - 1}{z - \alpha} = \frac{1 - Cz^n(\overline{\alpha}z - 1)}{z - \alpha}.$$

Clearly, P_n is a polynomial (of degree n) if and only if the last numerator in (2.9) vanishes at $z = \alpha$. We put therefore

$$(2.10) \qquad C := \frac{\alpha^{-n}}{|\alpha|^2 - 1}.$$

Theorem 2.7. *Let P_n and C be defined by (2.9) and (2.10). Then P_n is the polynomial of best approximation for f, and*

$$(2.11) \qquad E_n(f) = |C|.$$

Proof. For $|z| = 1$,

$$\left| \frac{\overline{\alpha}z - 1}{z - \alpha} \right| = \left| \frac{(\alpha/z) - 1}{z - \alpha} \right| = 1,$$

and this expression is less than 1 for $|z| < 1$ because $(\bar{\alpha}z - 1)/(z - \alpha)$ is an analytic function in $|z| \leq 1$. Let $F := f - P_n$ and let A_0 be the subset of A where the function $|F(z)|$ attains its maximum $\|F\|$. We see that A_0 is exactly the circumference $|z| = 1$, and that $\|F\| = |C|$.

It remains to show that for each polynomial Q of degree $\leq n$, (2.1) is satisfied, or what is the same thing, for some $|z| = 1$ and some integer k,

$$(2.12) \qquad \operatorname{Arg} F(z) - \operatorname{Arg} Q(z) = 2k\pi + \theta, \quad |\theta| \leq \pi/2.$$

For this purpose we study the behavior of $F(z)$ for $z = e^{it}$, $0 \leq t \leq 2\pi$. Since

$$\operatorname{Arg} F(z) = \text{const} + n \operatorname{Arg} z + \operatorname{Arg}(z - \overline{\alpha}^{-1}) - \operatorname{Arg}(z - \alpha),$$

this argument will increase by the amount $2\pi(n+1)$ as t changes from 0 to 2π. On the other hand, $\operatorname{Arg}(Q)$ increases less than $2\pi(n + \frac{1}{2})$ for otherwise the trigonometric polynomial $\operatorname{Re}(Q)$ would have more than $2n$ zeros. From this, we obtain (2.12), as desired. □

§ 3. Haar Systems

The finite dimensional subspaces X_n of $C(A)$ for which best approximation is unique have a special importance. They have many properties in common with the spaces of algebraic and trigonometric polynomials.

An important property of the functions $1, x, x^2, \dots, x^n$ is that a polynomial $a_0 + a_1 x + \cdots + a_n x^n$ that does not vanish identically can have no more than n zeros on an interval $[a, b]$. The functions $1, z, z^2, \dots, z^n$ enjoy the same property on each subset of the complex plane. Also, the trigonometric functions $1, \cos x, \sin x, \dots, \cos nx, \sin nx$ have this property on the circle $\mathbb{T}$. A trigonometric polynomial with real or complex coefficients

$$T(z) = a_0 + \sum_{k=1}^{n} (a_k \cos kz + b_k \sin kz) = \sum_{k=-n}^{n} c_k e^{ikz}$$

has modulo 2π at most $2n$ complex zeros (if z_0 is a zero of T then $z_0 + 2\pi k$, $k = 0, \pm 1, \dots$, are also zeros of T of the same multiplicity). For the proof, we write

$$(3.1) \qquad T(z) = \sum_{-n}^{n} c_k e^{ikz} =: e^{-inz} Q(w), \quad w = e^{iz}$$

where Q is an algebraic polynomial of degree $\leq 2n$. Now T has a zero at z_0 if and only if $w_0 = e^{iz_0} \neq 0$ is a zero of Q. By Leibniz' formula and induction, we see that $T(z_0) = \cdots = T^{(p)}(z_0) = 0$ is equivalent to $Q(w_0) = \ldots = Q^{(p)}(w_0) = 0$.

Hence z_0 has the same multiplicity for T as w_0 does for Q. Since Q has, counting multiplicity, at most $2n$ zeros $w_0 \neq 0$, T has at most $2n$ zeros modulo 2π.

We shall adopt the following definition: A set $\Phi = \{\phi_1, \dots, \phi_n\}$ of continuous, complex, or real functions on a Hausdorff topological space A is a *Haar system* if the following conditions are satisfied:

(a) A contains at least n points.
(b) Each $P = a_1\phi_1 + \cdots + a_n\phi_n$, which does not have all coefficients a_i equal to zero, has at most $n-1$ distinct zeros on A.

The space X_n spanned by $\phi_1, \dots, \phi_n$ is a *Haar space*. We call the elements of X_n polynomials (or Φ-polynomials). It is clear that any other basis $\Psi : \psi_1, \dots, \psi_n$ of X_n is also a Haar system. It follows in particular from this definition that the functions of a Haar system are linearly independent. The three systems mentioned at the beginning of this section furnish important examples of Haar systems. We leave it to the reader to prove that $1, \cos x, \dots, \cos nx$ is a Haar system on $[0, \pi]$, and $\sin x, \dots, \sin nx$ is a Haar system on $(0, \pi)$.

Condition (b) of the definition can be expressed also in the following form:

1. *If $x_1, \dots, x_n$ are distinct points of A, then the system of n equations with n unknowns $a_1, \dots, a_n$:*

$$a_1\phi_1(x_k) + a_2\phi_2(x_k) + \cdots + a_n\phi_n(x_k) = 0\,, \quad k = 1, \dots, n\,, \tag{3.2}$$

has only the obvious solution $a_1 = \cdots = a_n = 0$.

Using well-known facts about systems of linear equations, we see that this is also equivalent to each of properties **2** and **3**.

2. *If $x_1, \dots, x_n$ are distinct points of A, the determinant*

$$D(x_1, \dots, x_n) := D(\Phi;\ x_1, \dots, x_n) := \begin{vmatrix} \phi_1(x_1) & \dots & \phi_n(x_1) \\ \dots & \dots & \dots \\ \phi_1(x_n) & \dots & \phi_n(x_n) \end{vmatrix} \tag{3.3}$$

is not zero.

3. *If $x_1, \dots, x_n$ are distinct points of A and $c_1, \dots, c_n$ are arbitrary numbers, then the system of equations*

$$a_1\phi_1(x_k) + \cdots + a_n\phi_n(x_k) = c_k\,, \quad k = 1, \dots, n\,, \tag{3.4}$$

has a unique solution for the $a_1, \dots, a_n$.

If we put $P = a_1\phi_1 + \cdots + a_n\phi_n$, conditions (3.4) read $P(x_k) = c_k$, $k = 1, \dots, n$. We call this P an *interpolating polynomial* with prescribed valued c_k at the points x_k. Thus, statement **3** means that an interpolating polynomial P exists and is unique. If the number of the given points x_k and values c_k is less than n, then an interpolating polynomial also exists, but is not unique.

Here are some more examples of Haar systems:

Example 1. The functions $e^{a_i x}, a_1 < \cdots < a_n$ form a Haar system on any interval $[a, b]$.

This follows by induction on n; for $n = 1$ it is obvious. If $n > 1$ and $P(x) = \sum_1^n c_k e^{a_k x}$ has n zeros on $[a, b]$, then by Rolle's theorem $Q(x) = [e^{-a_1 x}P(x)]'$ has $n-1$ zeros. Since Q is a linear combination of $e^{(a_k - a_1)x}$, $k = 2, \dots, n$, Q must be identically zero; likewise P.

Example 2. A substitution of x for e^x shows that x^{a_i}, $a_1 < \cdots < a_n$ is a Haar system on any interval $[a, b]$, $a > 0$. One can also assume $a = 0$ here, if $a_1 = 0$.

Example 3. The functions $\Phi : 1, \dots, x^{n-1}, f(x)$ are a Haar system on $[a, b]$ if $f \in C^{n-1}$ and if $f^{(n)}$ exists and does not vanish on (a, b). Indeed, a polynomial in Φ has the form $P = P_{n-1} + \alpha f$, $P_{n-1} \in \mathcal{P}_{n-1}$. If P would vanish at $n+1$ distinct points, by Rolle's theorem, it would follow that $\alpha f^{(n)}(\xi) = 0$ for some $a < \xi < b$. Thus $\alpha = 0$ and then also $P_{n-1} = 0$.

In the remainder of this section, we shall consider real Haar systems. If a compact space A is given, does there exist a real Haar system for A? The answer is, in general, "no". We can prove that a set $A \subset \mathbb{R}^2$ carries no real Haar system with $n \geq 2$ if A contains three nonintersecting arcs emanating from a common points a. For example, there is no Haar system on a two-dimensional square.

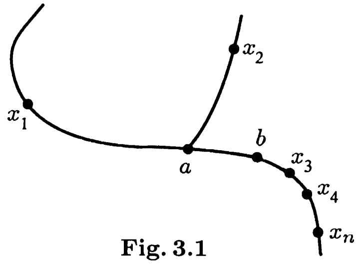


Fig. 3.1

In fact, assume that a Haar system exists for some $n \geq 2$; we select the points $x_1, \dots, x_n$ as indicated (Fig. 3.1). Put

$$g(x, y) := D(x, y, x_3, \dots, x_n)\,;$$

by **2**, $g(x, y) \neq 0$ as long as $x, y, x_3, \dots, x_n$ remain distinct. We move the points x, y continuously in A in the following manner: first, x moves from x_1 to b; then y moves from x_2 through a to x_1; finally, x moves from b through a to x_2. The function g will change continuously and remain different from zero; hence, $g(x_1, x_2)$ and $g(x_2, x_1)$ will have the same sign. But this is impossible, since $g(x_2, x_1) = -g(x_1, x_2)$.

We now assume that $\phi_1, \ldots, \phi_n$ is a real Haar system defined on $A = [a, b]$ or $A = \mathbb{T}$. It follows in this case by continuity that all of the determinants (3.3) for $x_1 < \cdots < x_n$ are of the *same sign.*

4. *If $x_1, \ldots, x_{n-1}$ are $n-1$ distinct points of A, then there exists a polynomial $D(x)$ that vanishes exactly at the points x_k and changes sign at each of these points (except when the point x_k coincides with a or b).*

If the ϕ_i are the powers $1, x, \ldots, x^{n-1}$ on $[a, b]$, we can simply take

$$D(x) := (x - x_1) \ldots (x - x_{n-1}) .$$

In the general case, we must modify this formula. We take (see (3.3))

$$D(x) := D(x, x_1, x_2, \ldots, x_{n-1}) . \tag{3.5}$$

Clearly, $D(x)$ vanishes for $x = x_1, \ldots, x_{n-1}$. If x is different from each of these points, $D(x) \neq 0$ according to **2.** Hence, D keeps a constant sign on each interval between any two points x_k. Let us show, for example, that D changes sign at x_1 (if x_1 is different from a, b). Take $h > 0$ so small that the interval $[x_1 - h, x_1 + h]$ is contained in A and does not contain the points $x_2, \ldots, x_{n-1}$. Consider the function:

$$g(t) := D(x_1 - h + t, x_1 + t, x_2, \ldots, x_{n-1}) ,$$

which is continuous and different from zero for $0 \le t \le h$. Clearly, g has a constant sign in this interval. But

$$g(0) = D(x_1 - h) , \quad g(h) = D(x_1, x_1 + h, x_2, \ldots, x_{n-1}) = -D(x_1 + h) .$$

This shows that $D(x_1 - h)$ and $D(x_1 + h)$ are of opposite sign and that D changes sign at x_1.

As a corollary of **4,** we note: *Each real Haar system on $\mathbb{T}$ consists of an odd number of functions.* Indeed, the continuous function D can have only an even number of sign changes on $\mathbb{T}$.

We wish to describe the sign changes of polynomials on $A = [a, b]$ or $A = \mathbb{T}$. In order to have identical proofs for these two cases, the following notation will be useful.

We shall say that a sequence of points $x_1, \ldots, x_p$ of $\mathbb{T}$ is ordered, $x_1 < \cdots < x_p (< x_1)$ if when moving in a positive direction we can go from x_1 to x_2, then from x_2 to x_3, and so on, without passing any of the previous points. We call disjoint closed subsets $A_1, \ldots, A_p$ of $A = [a, b]$ or of $A = \mathbb{T}$ *ordered on* A and write $A_1 < \cdots < A_p$ or $A_1 < \cdots < A_p < A_1$ respectively, if for any $x_1 \in A_1, \ldots, x_p \in A_p$, we have $x_1 < \cdots < x_p$.

5. *If $2 \le p \le n$, if p is even in case of $\mathbb{T}$, and if $A_1, \ldots, A_p$ are disjoint sets ordered on A, then there is a polynomial Q that does not vanish on the A_j and alternates in sign. In other words, for some $\varepsilon = \pm 1$, the sign of $Q(x)$ on A_j is $\varepsilon(-1)^j, j = 1, \ldots, p$. (The statement is also true for $p = 1$, see Theorem 9.2).*

By adding, if necessary, pairs of new sets in gaps between adjacent A_j, we reduce the statement to the cases $p = n$ and $p = n-1$. Let $A = \mathbb{T}$; the first case is impossible, since n is odd. In the second case, we select $n-1$ points x_i in the gaps between the A_i : $A_1 < \{x_1\} < A_2 < \cdots < \{x_{n-2}\} < A_{n-1} < \{x_{n-1}\} < A_1$ and take $Q := D$ of (3.5). Similarly, if $A = [a, b]$, $p = n$, we take $Q := D$ for points x_i with $A_1 < \{x_1\} < \cdots < \{x_{n-1}\} < A_n$. If $p = n - 1$, and $b \notin A_{n-1}$, we can take the points $A_1 < \{x_1\} < \cdots < \{x_{n-2}\} < A_{n-1} < \{x_{n-1}\} := \{b\}$. Similarly, if $a \notin A_1$. If $b \in A_{n-1}$, we can take the same D as before, but put $Q := D + \lambda Q_1$ where Q_1 is some polynomial with $Q_1(b) = 1$, and λ is small and of proper sign. □

In case $p \le n-2$ and $A = [a, b]$, $a \notin A_1$, $b \notin A_p$, the same proof yields a polynomial Q of the required type with $Q(a) = Q(b) = 0$.

For a Φ-polynomial P, we say that an interior point x_0 of A is a *double zero* of P if $P(x_0) = 0$ and P does not change sign in some neighborhood of x_0; all other zeros of P are *single* zeros.

6. *A non-trivial polynomial P can have at most $n-1$ zeros counting multiplicities, that is with double zeros counted twice.*

Let $\varepsilon > 0$ be so small that all zeros of P and the points a, b (if $A = [a, b]$) are at distance $> 2\varepsilon$ from each other. Let P have k_2 double zeros. The remaining zeros $x_1 < \cdots < x_{k_1}$ of P are zeros with change of sign or possible zeros at a, b (if $A = [a, b]$).

The complement in A of the union of the intervals $(x_i - \varepsilon, x_i + \varepsilon)$ for $i = 1, \ldots, k_1$ consists of p closed disjoint intervals $A_1 < \cdots < A_p$. On each A_j, P does not change sign, but the sets A_j contain all double zeros y of P. Since $k_1 \le n - 1$, we have $p = k_1 + 1 \le n$ sets A_j if none of a, b is a zero, $p = k_1 \le n-1$ sets if one of them is a zero, $p = k_1 - 1 \le n - 2$ if both are. In each of these cases, the proof of **5** or the remark after **5** produces a polynomial Q which on each A_j has the same sign as P and vanishes at a, b whenever P does. We put $Q_1 := P - \lambda Q$, where $\lambda > 0$ is sufficiently small. Then Q_1 has zeros a, b whenever P has them, changes sign on each interior interval $(x_i - \varepsilon, x_i + \varepsilon)$, and changes sign twice on each $(y - \varepsilon, y + \varepsilon)$. Since Q_1 has $\le n-1$ zeros, we obtain $2k_2 + k_1 \le n - 1$. □

§ 4. Uniqueness of Best Approximation in $C(A)$

Here we shall prove that for a Haar system $\Phi := \{\phi_1, \ldots, \phi_n\}$ of real or complex functions on a compact Hausdorff space A, each continuous function on A has only one polynomial of best approximation.

Lemma 4.1. *If A contains at least $n + 1$ points and P is a polynomial of best approximation to the continuous function f, then the set A_0 of all points $x \in A$, for which*

$$|f(x) - P(x)| = \|f - P\| =: E$$

contains at least $n + 1$ points.

Proof. Suppose that $A_0 = \{x_1, \dots, x_s\}$, where $s \le n$. Since there are points x with $|f(x) - P(x)| < E$, we must have $E > 0$. By **3** of §3, there is a polynomial Q such that

$$Q(x_k) = -[f(x_k) - P(x_k)], \quad k = 1, \dots, s.$$

Then

$$\max_{x \in A_0} \operatorname{Re}\left\{(f(x) - P(x))\overline{Q(x)}\right\} = \max_{1 \le k \le s}\{-|f(x_k) - P(x_k)|^2\} = -E^2 < 0$$

which contradicts Theorem 2.1. □

Theorem 4.2. *For a Haar system, there is a unique polynomial of best approximation for each continuous function.*

Proof. If A has exactly n points, then by **3** of §4, each function on A is equal to a polynomial P, and this polynomial is unique. Hence, we can assume that A consists of at least $n+1$ points; then Lemma 4.1 is applicable.

Assume that for a continuous function f there are two polynomials of best approximation, P and P_1:

$$\|f - P\| = \|f - P_1\| = E.$$

Then also $Q := \frac{1}{2}(P + P_1)$ is a polynomial of best approximation. By Lemma 4.1, there are at least $n+1$ points x for which

$$|f(x) - Q(x)| = E.$$

At each such point x, for the complex numbers

$$\alpha := f(x) - P(x), \quad \alpha_1 := f(x) - P_1(x),$$

we have $|\alpha + \alpha_1| = 2E$, $|\alpha| \le E$, $|\alpha_1| \le E$. But this is possible only if $\alpha = \alpha_1$. Thus, $P(x) = P_1(x)$ for at least $n+1$ points of A. Since Φ is a Haar system, the polynomials P and P_1 are identical. □

It is very remarkable that the converse of Theorem 4.2 is true.

Theorem 4.3 (A. Haar [1918]). *Let $\phi_1, \dots, \phi_n$ be linearly independent continuous functions defined on a compact Hausdorff space A that contains at least n points. If each continuous function has only one polynomial of best approximation, then $\phi_1, \dots, \phi_n$ is a Haar system.*

Proof. We shall assume that our system is not a Haar system, and construct a function f_1, which has several polynomials of best approximation.

By assumption, there exists n distinct points $x_1, \dots, x_n$ of A such that the system of linear equations

$$a_1\phi_1(x_k) + \cdots + a_n\phi_n(x_k) = 0, \quad k = 1, \dots, n \tag{4.1}$$

has nonzero solutions $a_1, \dots, a_n$. We select a solution of this kind and put

$$P_1(x) := a_1\phi_1(x) + \cdots + a_n\phi_n(x);$$

we can assume that $\|P_1\| = 1$.

The transposed system

$$c_1\phi_i(x_1) + \cdots + c_n\phi_i(x_n) = 0, \quad i = 1, \dots, n, \tag{4.2}$$

also has nonzero solutions $c_1, \dots, c_n$. Let $c_1, \dots, c_n$ be such a solution normalized so that $\sum_1^n |c_i| = 1$. Some of the c_k may be zero, but the set Λ of k for which $c_k \ne 0$ is not empty. The formula $\lambda_0(f) = \sum_1^n c_k f(x_k)$ defines a continuous linear functional on $C(A)$, and $\|\lambda_0\| = \Sigma|c_k| = 1$. For each ϕ_i, $\lambda_0(\phi_i) = 0$, hence λ_0 is orthogonal to all polynomials.

According to the Tietze theorem about the extension of continuous functions, there is a function $f_0 \in C(A)$ with the properties:

$$\begin{cases} f_0(x_k) = \bar{c}_k/|c_k|, & k \in \Lambda \\ |f_0(x)| \le 1, & x \in A. \end{cases} \tag{4.3}$$

We put

$$f_1(x) := (1 - |P_1(x)|)f_0(x).$$

We use the remark (1.7) to show that αP_1 is a best approximation for f_1 for each $0 \le \alpha \le 1$. Since $\lambda_0(f_1) = \Sigma c_k f_0(x_k) = 1$, the degree of approximation of f_1 by (1.7) is ≥ 1. It is exactly 1, and all αP_1 are best approximations because for all $x \in A$,

$$|f_1(x) - \alpha P_1(x)| \le |f_1(x)| + \alpha|P_1(x)| \le 1 - |P_1(x)| + |P_1(x)| \le 1.$$ □

§5. Chebyshev's Theorem

For real Haar systems $\Phi : \phi_1, \dots, \phi_n, n \ge 2$, on $A = [a,b]$ or $\mathbb{T}$ there is a much simpler characterization of best approximants than that given by Theorem 2.2. We shall discuss best approximation on any closed subset B of A which contains at least $n+1$ points. Let $P^* \ne f$ be the Φ-polynomial of best approximation to f in $C(B)$.

We state first some properties of the approximation functionals λ^* of (2.8) which satisfy Corollary 2.6. Such functionals have the form

$$\lambda^*(g) := \sum_{i=0}^{r} \alpha_i g(x_i), \quad \sum_{i=0}^{r} |\alpha_i| = 1, \quad g \in C(B) \tag{5.1}$$

and we have $r \le n$ and $\lambda^*(\phi_j) = 0$, $j = 1, \dots, n$. We order the points $x_0 < \cdots < x_r$ in the positive direction on A.

1. *One has $r = n$ and $\alpha_i \ne 0$ for all i.*
2. *The values of α_i alternate in sign.*

Indeed, if **1** were not valid then by property **3** of §3, there would be a polynomial Q whose sign at x_i agrees with that of the $\alpha_i, i = 0, \ldots, r$, and we would obtain the contradiction $\lambda^*(Q) > 0$. Similarly, if **2** did not hold we could take disjoint intervals $A_1, \ldots, A_p$, $p \le n$, ordered in the positive direction on A, containing the x_j and such that for all $x_i \in A_j$, all of the corresponding α_i have the same sign, $j = 1, \ldots, p$. By **5** of §3, there would exist a polynomial Q which has at the x_i the same sign as α_i for $i = 0, \ldots, n$, and we obtain the contradiction $\lambda^*(Q) > 0$.

For any polynomial P, let $d := \|f - P\|(B)$ and let $B_0 := \{x \in B : |f(x) - P(x)| = d\}$.

Theorem 5.1. *Let $\Phi = \{\phi_1, \ldots, \phi_n\}$ be a real Haar system on A, where $A = [a, b]$ or $\mathbb{T}$, and let $B \subset A$ have at least $n+1$ points. If $f \in C(B)$, then a polynomial P is the best approximation to f on B if and only if there are $n+1$ points $x_0 < x_1 < \cdots < x_n$ in B_0 and a number $\varepsilon = \pm 1$ for which*

$$f(x_i) - P(x_i) = \varepsilon(-1)^i d, \quad i = 0, \ldots, n. \tag{5.2}$$

Proof. Let P^* be the best approximation to f on B and let λ^* be a best approximation functional for f and P^*. From the properties (2.8), $\lambda^*(f - P^*) = \|f - P^*\|$, and $f(x_i) - P^*(x_i) = \|f - P^*\| \operatorname{sign} \alpha_i$, $i = 0, \ldots, n$, which is (5.2). Conversely, suppose that P satisfies (5.2) and Q is any polynomial. Since Q has at most $(n-1)$ changes of sign, for one of the points $x_0, \ldots, x_n$ we have $\operatorname{sign} Q(x_i) = \varepsilon(-1)^i$ and hence from Theorem 2.2, P is a best approximation to f. □

The points x_i in (5.2) are called *alternation points* for the approximating polynomial P. In particular, an algebraic polynomial of degree $\le n$ of best approximation to f must have at least $n+2$ alternation points. Similarly, a trigonometric polynomial of degree $\le n$ of best approximation has at least $2n+2$ alteration points.

Because of (5.2), a *polynomial P of best approximation* on $A = [a, b]$ or $\mathbb{T}$ *interpolates* f : there are at least n distinct interior points $\xi_i \in A$ with $P(\xi_i) = f(\xi_i)$, $i = 1, \ldots, n$, in case of $A = [a, b]$, at least $n+1$ points of $A = \mathbb{T}$.

If B consists of $n+1$ points $x_0, \ldots, x_n$, Theorem 5.1 shows that the polynomial P of best approximation to f on B is determined by the system of equations

$$f(x_i) - P(x_i) = (-1)^i d, \quad i = 0, \ldots, n, \tag{5.3}$$

whose unknowns are d and the coefficients of $\phi_1, \ldots, \phi_n$. Then $d = \pm\|f - P\|(B)$.

A useful application of the alternation theorem is the following lower estimate for the degree of approximation.

Theorem 5.2 (de la Vallée Poussin [A-1919, p. 85]). *Let Φ be a Haar system on $A = [a, b]$ or $\mathbb{T}$ and let B be a closed subset of A containing at least $n+1$ points. If for a Φ-polynomial P, there is an ordered set of $n+1$ distinct points x_i from B for which $f(x_i) - P(x_i) = \varepsilon(-1)^i \mu_i$, with $\varepsilon = \pm 1$ and $\mu_i > 0$, $i = 0, \ldots, n$ then $E_n(f) \ge \min_i \mu_i$.*

Proof. If we have $E_n(f) < \mu_i$, for all i, then for the best approximant P^* to f, we have $\operatorname{sign}[P^*(x_i) - P(x_i)] = \operatorname{sign}[(f(x_i) - P(x_i)) - (f(x_i) - P^*(x_i))] = \varepsilon(-1)^i$. Hence, $P - P^*$ has n sign alternations and we have the desired contradiction $P = P^*$. □

§ 6. Chebyshev Polynomials

Cases when there is an explicit formula for the degree of approximation $E_n(f)$, or for the polynomial of best approximation for f, are exceptional and are of special interest. We shall give some examples.

We begin with the approximation of x^n on $[-1, +1]$ by polynomials of degree $\le n-1$ in the uniform norm. We wish to find a polynomial $a_{n-1}x^{n-1} + \cdots + a_1 x + a_0$ with real coefficients such that

$$\max_{-1 \le x \le 1} \left| x^n - a_{n-1}x^{n-1} - \cdots - a_0 \right|$$

takes its smallest possible value. Clearly, this is equivalent to the following problem: among all polynomials $P(x) = x^n + a_{n-1}x^{n-1} + \cdots + a_0$ of degree n with leading coefficient 1, find the polynomial of least deviation from zero, that is, a P with smallest possible norm.

Since $1, x, \ldots, x^{n-1}$ is a Haar system on $[-1, +1]$, Theorem 5.1 is applicable, and so we conclude that a polynomial P_n is a polynomial of least deviation from zero, if and only if (a) there exist $(n+1)$ points of $[-1, +1]$ where P_n takes the values $\pm\|P_n\|$ with alternating signs, and (b) P_n has leading coefficient 1.

It is easy to find a P_n satisfying (a). For each x in $[-1, 1]$, there is a unique t, $0 \le t \le \pi$, with $x = \cos t$. Consequently, $C_n(x) := \cos nt$, $n = 0, 1, \ldots$, are functions of x. We have $C_0(x) = 1$, $C_1(x) = x$. The C_n satisfy the recurrence relation

$$C_n(x) = 2xC_{n-1}(x) - C_{n-2}(x), \quad n = 2, 3, \ldots, \tag{6.1}$$

which follows from

$$\cos nt + \cos(n-2)t = 2\cos t \cos(n-1)t .$$

We compute easily:

$$C_2(x) = 2x^2 - 1, \quad C_3(x) = 4x^3 - 3x, \quad C_4(x) = 8x^4 - 8x^2 + 1, \ldots .$$

We see that C_n, $n = 0, 1, \ldots$, is an algebraic polynomial of degree n with leading coefficient 2^{n-1}, $n \ge 1$. This follows from (6.1) by induction.

Theorem 6.1. *For $n = 1, 2, \ldots$, the polynomial of degree n with leading coefficient 1 of least deviation from zero on $[-1, +1]$ is $2^{-n+1}C_n(x)$. The degree of approximation of x^n on $[-1, +1]$ by polynomials of degree $n-1$ is 2^{-n+1}.*

Proof. We have $\|C_n\| = \max_t |\cos nt| = 1$. Also, $\cos nt = (-1)^k$ for $t = k\pi/n$, $k = 0, \dots, n$; hence, $C_n(x_k) = (-1)^k$, where $x_k = \cos(k\pi/n)$, $k = 0, \dots, n$. Thus, C_n satisfies (a). □

The polynomials C_n are called the *Chebyshev polynomials.* We note some of their remarkable properties.

1. The explicit expression for $C_n(x)$, valid for all complex x is

$$C_n(x) = \frac{1}{2}\left\{\left(x + \sqrt{x^2-1}\right)^n + \left(x - \sqrt{x^2-1}\right)^n\right\}. \tag{6.2}$$

Adding the two formulas

$$\cos nt \pm i \sin nt = (\cos t \pm i \sin t)^n,$$

we obtain

$$\cos nt = \tfrac{1}{2}\left((\cos t + i \sin t)^n + (\cos t - i \sin t)^n\right).$$

From this, (6.2) follows, for if $\cos t = x$, then $i \sin t$ is one of the two values of $\pm\sqrt{x^2-1}$. This is (6.2) for $-1 \le x \le 1$. Since both sides of (6.2) are polynomials, this relation holds for all x.

2. The C_n, $n = 0, 1, \dots$, form a system of *orthogonal polynomials* on $[-1, +1]$ with weight-function $1/\sqrt{1-x^2}$:

$$\int_{-1}^{+1} \frac{1}{\sqrt{1-x^2}} C_n(x) C_m(x)\, dx = 0, \quad n \ne m. \tag{6.3}$$

This follows from $\int_0^\pi \cos nt \cos mt\, dt = 0$, $n \ne m$.

3. The *zeros* of C_n can be easily found. Since $\cos nt = 0$ if $t = (2k-1)\pi/(2n)$,

$$x_k = \cos\frac{2k-1}{2n}\pi, \quad k = 1, \dots, n \tag{6.4}$$

are the n zeros of C_n. Thus, the zeros are always real, distinct, and lie in $(-1+1)$. They are symmetric with respect to the point 0. The points $\exp(it_k)$ are equally spaced on the upper arc of the unit circle. As a result, the x_k are more densely distributed around ± 1 than in the interior of $(-1, +1)$.

4. Differentiating the relation $C_n(x) = y = \cos(n \arccos x)$, we obtain $\sqrt{1-x^2}y' = n \sin(n \arccos x)$ and a second differentiation yields *the differential equation of the Chebyshev polynomials.*

$$(1-x^2)y'' - xy' + n^2 y = 0. \tag{6.5}$$

See the book of Rivlin [A-1990] for more information about Chebyshev polynomials.

§ 7. Strong Unicity

If for each f in the Banach space X, best approximation to f from X_n is unique, then by Theorem 1.2, the best approximant Pf is a continuous function of f:

$$\|Pf - Pf_0\| \to 0 \quad \text{if} \quad \|f - f_0\| \to 0. \tag{7.1}$$

However, in general, the operator P is not linear. For example, for approximation by the elements of $\mathcal{P}_n$ in $C[0, \pi]$, let $f(x) := \sin 2nx$, $x \in [0, \pi]$ and let g be derived from f by replacing f by zero between the last two zeros of f. By Chebyshev's Theorem, for $n \ge 3$, $Pf = Pg$ and $E_n(f) = E_n(g) = 1$. But the zero polynomial is not the best approximation to $f - g$ since the polynomial $Q \equiv 1/2$ approximates this difference with error $1/2$.

We shall try to go further than the continuity of P. We assume that $X = C(A)$ with $A = [a, b]$ or $\mathbb{T}$ and that $\Phi := \{\phi_1, \dots, \phi_n\}$ is a Haar system on A. Then we have the following two theorems.

Theorem 7.1 (Newman-Shapiro [1963]). *The operator of best approximation P is strongly unique: for each $f_0 \in C(A)$ there is a constant $\gamma > 0$ with the property*

$$\|f_0 - Q\| \ge \|f_0 - Pf_0\| + \gamma\|Q - Pf_0\| \tag{7.2}$$

for any polynomial Q.

Inequality (7.2) is stronger than the uniqueness property of Pf, which means that $\|f_0 - Q\| > \|f_0 - Pf_0\|$, if $Q \ne Pf_0$.

Theorem 7.2 (Freud [1958]). *The operator P satisfies locally the Lipshitz condition: for each $f_0 \in C(A)$ there exists a constant $\delta > 0$ for which*

$$\|Pf - Pf_0\| \le \delta\|f - f_0\|, \quad f \in C(A). \tag{7.3}$$

First we show that (7.2) implies (7.3) with $\delta = 2/\gamma$. Letting $Q := Pf$ in (7.2), we get

$$\begin{aligned}
\gamma\|Pf - Pf_0\| &\le \|f_0 - Pf\| - \|f_0 - Pf_0\| \\
&\le \|f - f_0\| + \|f - Pf\| - \|f_0 - Pf_0\| \\
&\le \|f - f_0\| + \|f - Pf_0\| - \|f_0 - Pf_0\| \\
&\le 2\|f - f_0\|.
\end{aligned}$$

Proof of Theorem 7.1. We can assume that $f_0 \ne Pf_0$. We let $\lambda^*(g) := \sum_{i=0}^r \alpha_i g(x_i)$ be a linear functional (2.8) for the function f_0 and the polynomial $P := Pf_0$. We have shown in §5 that $r = n$ and $\sigma_i := \operatorname{sign}\alpha_i = \operatorname{sign}(f_0(x_i) - Pf_0(x_i))$ alternate in sign. For any Φ-polynomial Q and any i,

$$\begin{aligned}\|f_0 - Q\| &\geq \sigma_i(f_0 - Q)(x_i) = \sigma_i(f_0 - Pf_0)(x_i) + \sigma_i(Pf_0 - Q)(x_i). \\ &= \|f - Pf_0\| + \sigma_i(Pf_0 - Q)(x_i).\end{aligned} \tag{7.4}$$

We take a maximum over all i in (7.4) and obtain

$$\|f_0 - Q\| \geq \|f_0 - Pf_0\| + \max_{0\leq i\leq n} \sigma_i(Pf_0 - Q)(x_i). \tag{7.5}$$

Now, for the given σ_i, $\Lambda(P) := \max_{0\leq i\leq n} \sigma_i P(x_i)$ is a continuous function of the Φ-polynomial P. Let $\|P\| = 1$. Since the σ_i alternate in sign and P has at most $n-1$ zeros, one of the $\sigma_i P(x_i)$ is positive. It follows that for some constant $\gamma > 0$, we have $\Lambda(P) \geq \gamma$, for all P of norm one. Moreover, Λ is homogeneous: $\Lambda(aP) = a\Lambda(P)$ if $a > 0$ and therefore $\Lambda(P) \geq \gamma\|P\|$ for all P. If we use this in (7.5) for $P = Pf_0 - Q$, we obtain (7.2). □

§ 8. Remez Algorithms

The Chebyshev alternation theorem suggests ways for numerically computing the best uniform approximant P^* to $f \in C[a,b]$ by polynomials from a Haar system $\Phi := \{\phi_1, \ldots, \phi_n\}$. The Remez algorithm (Remez [1934]) consists of the following. We select a sequence of $n+1$ points $T : a \leq t_0 < t_1 < \cdots < t_n \leq b$ and approximate P^* by the polynomial $P = P_T = \sum_1^n a_j\phi_j$ which approximates f best on the set T. According to (5.3), we can find P from the system of equations:

$$\sum_{j=1}^n a_j\phi_j(t_i) + (-1)^i d = f(t_i), \quad i = 0, 1, \ldots, n \tag{8.1}$$

with unknowns a_j, $j = 1, \ldots, n$, and d. It follows that $|d|$ is the error of best approximation to f on T. The de la Vallée Poussin theorem (Theorem 6.2) ensures that the error of approximation $E(f)$ on $A := [a,b]$ satisfies

$$|d| \leq E(f) \leq D, \quad D := \|f - P\|(A). \tag{8.2}$$

The goal is to choose the sequence T so that $\Delta := D - |d|$ is small. This guarantees that P is close to P^*. Indeed, from Theorem 7.1,

$$\|P - P^*\| \leq \gamma^{-1}[\|f - P\| - \|f - P^*\|] \leq \gamma^{-1}\Delta, \tag{8.3}$$

with the constant γ depending only on f.

For a Remez algorithm, sequences of $n+1$ points $T_0, T_1, \ldots$ and best approximations $P_0, P_1, \ldots$ are defined recursively. The sequence T_{k+1} is determined from T_k by changing some of the points in T_k (with the effect of increasing the size of $|d_k|$). We shall only consider the *single exchange algorithm* where exactly one point is changed at each step. Let T denote one of the sequences T_k and let T_+ denote the next updated sequence T_{k+1}; we use similar notation for the updated values of d, D and P. We let ξ be a point (to be found numerically) where $|f(\xi) - P(\xi)| = D$. The sequence T_+ is obtained from T by replacing one of the points t_h by ξ and leaving all other points unchanged. We choose h so that $f - P$ alternates in sign at the points of the new sequence $T_+ : t_0^+ < \cdots < t_n^+$.

To prove the convergence of the Remez algorithm, we estimate how much d is increased by a single step. Let λ_+^* be an approximation functional (2.8) for f and P_+ and the set T_+. Since T_+ has just $n+1$ points, the x_i of (2.8) must be the t_i^+, so that $\lambda_+^*(g) =:= \sum_0^n \alpha_i^+ g(t_i^+)$. We have $|f(t_i^+) - P_+(t_i^+)| = |d_+|$, $i = 0, \ldots, n$. From the properties of the α_i of §5, we have

$$\begin{aligned}|d_+| &= \lambda_+(f) = \lambda_+(f - P) \\ &= |d|\sum_{i=0}^n |\alpha_i^+| + |\alpha_h^+|(D - |d|) = |d| + |\alpha_h^+|(D - |d|).\end{aligned} \tag{8.4}$$

Unless $|d| = D$, in which case $P_k = P^*$ and the process stops, we have $|d_k| < |d_{k+1}|$. We also see that if $f \neq P^*$, then $q := |d_1| > 0$. To apply (8.4), we shall need the following lemma.

Lemma 8.1. *In the above situation, for $k \geq 1$, one has $|d| \geq q$ and moreover*

$$t_i - t_{i-1} \geq C_1, \quad i = 1, \ldots, n \tag{8.5}$$

$$|\alpha_i| \geq C_2, \quad i = 0, \ldots, n. \tag{8.6}$$

where the constants $C_1, C_2 > 0$ depend only on q, f and Φ.

Proof. Since $q := |d_1|$, it follows by induction from (8.4) that $d \geq q$. Let ω be a modulus of continuity for which $\omega(f,t) \leq \omega(t)$ and $\omega(\phi_k, t) \leq \omega(t)$, $t > 0$, $k = 1, \ldots, n$. Since any two norms on the finite dimensional space X_n are equivalent, for some constant $C_0 > 0$, and any $Q = \sum_{k=1}^n a_k\phi_k$, we have $\sum_{k=1}^n |a_k| \leq C_0\|Q\|$. We apply this to the polynomial P of best approximation to f on T and obtain for $i = 1, \ldots, n$,

$$2q \leq 2d = |[f(t_i) - P(t_i)] - [f(t_{i-1}) - P(t_{i-1})]| \leq (1 + C_0\|P\|)\omega(t_i - t_{i-1}).$$

Then (8.5) follows by choosing C_1 so small that $\omega(C_1) \leq 2q/(1 + C_0\|P\|)$.

To prove (8.6), we note that the determinant $D(y_1, \ldots, y_n)$ of (3.3) is continuous and nonzero on the closed set $\{(y_1, \ldots, y_n) : |y_i - y_{i-1}| \geq C_1,\ |y_i| \leq 1\}$. In particular, the determinants $\delta_i := D(t_0, \ldots, t_{i-1},\ t_{i+1}, \ldots, t_n)$, $i = 0, \ldots, n$, are all of the same sign and satisfy $|\delta_i| \geq C$ for a constant $C > 0$ depending only on f and Φ. Now, the coefficients α_i of the approximation functional λ_T of (2.8) for f, P and the set T are a solution of the system of equations

$$\begin{cases} \sum\limits_{i=0}^n (-1)^i\alpha_i = \varepsilon \\ \sum\limits_{i=0}^n \alpha_i\phi_j(t_i) = 0, \quad j = 1, \ldots, n \end{cases} \tag{8.7}$$

where $\varepsilon = \pm 1$. The coefficient matrix of (8.7) has determinant $\widetilde{\delta} := \sum_{k=0}^{n} \delta_k$. Since $|\widetilde{\delta}|$ is bounded from above by a constant depending only on Φ and $|\alpha_i| = |\delta_i / \widetilde{\delta}|$, (8.6) follows. □

If we use (8.6) in (8.4), we obtain

$$|d_+| \ge |d| + C_2(D - |d|) . \tag{8.8}$$

From this, we can easily derive the convergence of the one point exchange algorithm.

Theorem 8.2. *If T_0 is a set of $n+1$ points for which $|d_0| > 0$, then the Remez one point exchange algorithm converges: $|d_k| \to E(f)$ and $P_k \to P^*$ as $k \to \infty$.*

Proof. We have by (8.8), $E(f) \ge |d_{k+1}| \ge |d_k| + C_2(D_k - |d_k|)$. Since $D_k \ge E(f)$, this gives

$$\begin{aligned} E(f) - |d_{k+1}| &\le E(f) - |d_k| - C_2(E(f) - |d_k|) \\ &= (1 - C_2)(E(f) - |d_k|) . \end{aligned} \tag{8.9}$$

Since $C_2 > 0$, we have $|d_k| \to E(f)$. Moreover, from (8.8), $D_k - |d_k| \le C_2^{-1}(|d_{k+1}| - |d_k|)$, and therefore (8.3) gives $P_k \to P^*$. □

We note that the convergence $|d_k| \to E(f)$ and the uniform convergence of P_k to P^* are geometric: for some $0 < \rho < 1$, and $C > 0$, $\|P^* - P_k\| \le C\rho^k$.

§ 9. Krein's Theorem

In some applications it is important to know to what extent the zeros of a polynomial P from a Haar system $\Phi := \{\phi_1, \ldots, \phi_n\}$ can be prescribed.

Theorem 9.1 (Krein [1951]). *Let Φ be a Haar system on $A = [a,b]$ or $A = \mathbb{T}$. Let $x_1, \ldots, x_{k_1}, y_1, \ldots, y_{k_2}$ be distinct points of A, with neither a nor b among the y_k if $A = [a,b]$. If $k_1 + 2k_2 \le n - 1$, and if k_1 is even in case $A = \mathbb{T}$, then there exists a polynomial Q with simple zeros at x_i, double zeros at y_j, and with no other zeros on A. The only exceptional case is when $n - k_1 - 1$ is odd (then necessarily $A = [a,b]$) and exactly one of the x_i is an endpoint. Then Q may have an additional zero at the other endpoint.*

An important special case of Krein's theorem is:

Theorem 9.2. *There is a polynomial Q that is strictly positive on A.*

Proof of Theorem 9.1. Let $n - 1 = k_1 + 2k_2 + 2l + \varepsilon$ with $\varepsilon = 0$ or 1. We introduce l interior points z_k distinct from the x_i, y_j. We take m so large that the points

$$x_i\,, \quad y_j - 1/m\,, \quad y_j + 1/m\,, \quad z_k - 1/m\,, \quad z_k + 1/m \tag{9.1}$$

are distinct and contained in A.

Assume first that $\varepsilon = 0$, then (9.1) consists of $n-1$ points. By **4** of §3, there is a polynomial P_m which vanishes at these points, changes sign at each of them unless they are endpoints a, b, and satisfies $\|P_m\|_\infty = 1$. An appropriate subsequence P_{m_j} converges to a polynomial P. Then P has single zeros at the x_i, double zeros at the y_j and z_k and no other zeros (because they total $n-1$). Taking points z_k^* different from the z_k, we obtain polynomials P_m^* and P^*. We can also suppose $\operatorname{sign} P^* = \operatorname{sign} P$ at all non-zero points. The polynomial $Q = P + P^*$ is the desired polynomial.

If $\varepsilon = 1$, then $A = [a,b]$. We modify the above construction by adding an additional point w to those in (9.1). If neither endpoint appears among the x_i then w is taken as a in the definition of P_m and as b in the definition of P_m^*; the resulting polynomial $Q = P + P^*$ will not have a zero at either endpoint. If only one endpoint is an x_i, then we take w as the other endpoint in the construction of both P_m and P_m^*. In this case, Q has a zero at both endpoints; that is, Q has then one undesirable (endpoint) zero. Finally, if both endpoints appear among the x_i, then we take $w = a + \frac{1}{m}$ in the construction of both P_m and P_m^*. In this case, P will have zeros only at the x_i, y_i, z_k since an additional (interior) zero of P would necessarily be a double zero and the total number of zeros would exceed $n-1$. Similarly for P^*. In this case Q has zeros x_i, y_k only. □

Often the exceptional case in Theorem 9.1 cannot occur. This happens for example if Φ has an extension (as a Haar system) to a larger interval $[a', b']$, $a' < a < b < b'$. But the exceptional case cannot be completely eliminated.

Example (Haverkamp [1978]). Let Φ_0 be the system

(9.2)
$$\phi_1 := 1\,, \quad \phi_2 := x^3 - 2x\,, \quad \phi_k := x^{k-3}(x^2-1)^2\,, \quad k = 3, \ldots, n, \text{ on } I = [-1,1]\,.$$

For odd $n \ge 3$, (9.2) is a Haar system. For this system, additional zeros in Theorem 9.1 at the endpoints cannot be eliminated.

Proof. We assume that $P = c_1 + c_2\phi_2 + \cdots + c_n\phi_n$ is a non-trivial polynomial with n distinct zeros in $[-1,1]$ and derive a contradiction. It is sufficient to consider the cases: 1) $c_n = 1$ and 2) $c_n = 0$, $c_{n-1} \ne 0$, since otherwise P is an algebraic polynomial of degree less than n.

We can also assume that $c_2 \le 0$. In case 2), we can replace P by $-P$ if necessary. In case 1), ϕ_n is even, and we use either $P(x)$ or $P(-x) = c_1 - c_2\phi_2(x) + \cdots + c_n\phi_n(x)$.

Case 1). From $\lim_{x\to\pm\infty} P(x) = +\infty$, we see that P has an even number of zeros, hence $n+1$ zeros. These zeros cannot all be in $(-\infty, 1]$, because Rolle's theorem would then give n zeros of P' in $(-\infty, 1)$. In addition, the relations $P'(1) = c_2 \le 0$ and $\lim_{x\to\infty} P'(x) = +\infty$ imply that P' has a zero on $[1,\infty)$. Hence, altogether $n+1$ zeros in $(-\infty,+\infty)$. But then, $P' \equiv 0$ would imply $P \equiv 0$.

Thus P has exactly one simple zero on $(1,\infty)$. But this can happen only if $P(1) = c_1 - c_2 < 0$. On the other hand, in this case, P cannot have a zero on $(-\infty,-1)$, hence $P(-1) = c_1 + c_2 \ge 0$. These two inequalities contradict $c_2 \le 0$.

Case 2). Here P is of degree n. We have $c_2 \ne 0$, for otherwise P' would be divisible by x^2-1, hence by Rolle's theorem, it would have $n+1$ zeros in $[-1,+1]$, leading to $P' \equiv 0$.

Returning to the proof, we note that all $(n-1)$ zeros of P' are in $(-1,+1)$. Since $P'(\pm 1) = c_2 < 0$, we have $\lim_{x\to\pm\infty} P'(x) = -\infty$ and therefore $\lim_{x\to-\infty} P(x) = +\infty$, $\lim_{x\to+\infty} P(x) = -\infty$. This leads to $P(-1) = c_1+c_2 \ge 0$, $P(1) = c_1 - c_2 \le 0$, hence to $c_2 \ge 0$, a contradiction.

For the system Φ_0, there is no polynomial P which vanishes only at -1. Indeed, we can assume that $P(1) = c_1 - c_2 > 0$. But then $P(-1) = c_1 + c_2 = 0$ implies $P'(-1) = c_2 < 0$. It would follow that P takes negative values near -1 and hence P has another zero on $(-1,1)$. □

The system Φ_0 is remarkable also in another respect. *It is not extendable as a Haar system onto any larger interval* $J \supset I$. For example, if Φ_0 could be extended onto $[-1,a]$, $a>1$, then by Theorem 9.1, there would exist in the extended system a P with $P(-1) = P(a) = 0$, $P(x) > 0$ on $(-1,a)$.

§ 10. Best Approximation in L_p, $1 \le p < \infty$

On can find characterization theorems for best approximation in the spaces L_p, $1 \le p < \infty$, by functional analytic means, deriving them from Theorem 1.3. They are valid in more general spaces. For a Banach space X, and $f \in X$, $f \ne 0$, there are always linear functionals λ^* satisfying

$$\|\lambda^*\| = 1\,, \quad \lambda^*(f) = \|f\|\,. \tag{10.1}$$

We shall say that X has a *smooth unit ball,* or simply that X *is smooth,* if for each $f \in X$, λ^* is unique; we then write $\lambda_f := \lambda^*$.

Theorem 10.1. *Let $f \in X\backslash Y$ and $g_0 \in Y$ where Y is a closed linear subspace of a smooth Banach space X. Then g_0 is a best approximation to f if and only if*

$$\lambda^*(g) = 0 \textit{ for all } g \in Y\,, \quad \lambda^* := \lambda_{f-g_0}\,. \tag{10.2}$$

Proof. If g_0 is a best approximation to f, let λ_0 be the functional of Theorem 1.3. Then λ_0 is orthogonal to Y, and therefore $\|f-g_0\| = \lambda_0(f-g_0)$. We must have $\lambda_0 = \lambda^*$. It follows that λ^* is orthogonal to Y.

Conversely, let (10.2) be satisfied. Then for all $g \in Y$,

$$\|f-g_0\| = \lambda^*(f-g_0) = \lambda^*(f-g) \le \|f-g\|\,. \qquad \square$$

For example, the space $L_p(A,d\mu)$ with a positive, σ-finite, measure $d\mu$ is smooth if $1 < p < +\infty$. Indeed, the uniqueness of $\lambda^* = \lambda_f$ in (10.1) follows easily from the Riesz representation theorem and the case of equality in Hölder's inequality. In fact,

$$\lambda_f(g) = \int_A g h\, d\mu\,, \quad h := \frac{|f|^{p-1}\operatorname{sign} f}{\|f\|^{p-1}} \in L_{p'}\,, \quad 1/p + 1/p' = 1\,. \tag{10.3}$$

As usual, for real c, sign(c) is defined to be 0 if $c = 0$.

Corollary 10.2. (i) *If $f \in L_p\backslash Y$, $1 < p < \infty$, then $g_0 \in Y$ is a best approximation to f if and only if the function*

$$h := |f-g_0|^{p-1}\operatorname{sign}(f-g_0) \tag{10.4}$$

is orthogonal to Y. (ii) *If $f \in L_1\backslash Y$, $g_0 \in Y$ is a best approximation to f if and only if there exists a function h, $|h(t)| \le 1$* a.e., *which on the set $B := \{t \in A : f(t) \ne g_0(t)\}$ is equal to* sign$(f-g_0)$, *and which is orthogonal to Y.*

Proof. (i) is a consequence of Theorem 10.1. For (ii), we appeal to Theorem 1.3, and see that the conditions of this theorem are equivalent to the existence of a function h for which (a) esssup $|h(t)| = 1$, (b) $\int_A |f-g_0|\,d\mu = \int_A (f-g_0) h\, d\mu$ and (c) h is orthogonal to Y. □

Condition (ii) takes a particular simple form if $f(t) = g_0(t)$ only on a set of measure zero. Then *the condition is simply that* $h = \operatorname{sign}(f-g_0)$ *is orthogonal to Y.*

In applications of Corollary 10.2 the following fact is often useful:

Lemma 10.3. *If for some positive integer k the function $g \in L_1(\mathbb{T})$ has period $2\pi/k$, then for any l which is not divisible by k,*

$$\int_{\mathbb{T}} g(t)\cos lt\,dt = \int_{\mathbb{T}} g(t)\sin lt\,dt = 0\,. \tag{10.5}$$

Proof. In fact,

$$\int_0^{2\pi} g(t)e^{ilt}dt = \int_{2\pi/k}^{2\pi+2\pi/k} g(t)e^{ilt}dt = e^{\frac{2\pi i l}{k}} \int_0^{2\pi} g(t)e^{ilt}dt\,. \qquad \square$$

For example, suppose that g has period $2\pi/(n+1)$ on $\mathbb{T}$ and that $\int_{\mathbb{T}} |g|^{p-1} \operatorname{sign} g\, dt = 0$, then $|g|^{p-1} \operatorname{sign} g$ is orthogonal to all trigonometric polynomials of degree $\le n$. Corollary 3.2 yields then that the zero function is a best approximation to g from $\mathcal{T}_n$, $1 \le p < \infty$.

It is possible to obtain a reformulation of Corollary 10.2 (ii) by means of the functions f and g_0 alone. We define $A_0 := A_0(f, g_0) := \{x : f(x) - g_0(x) = 0\}$. We have the following theorem of Walsh and Motzkin [1959].

Theorem 10.4. *For a subspace Y of $L_1(A, d\mu)$, the function $g_0 \in Y$ is a best approximation to $f \in L_1$ from Y if and only if for all $g \in Y$,*

$$\int_A g \operatorname{sign}(f - g_0) d\mu \le \int_{A_0} |g| d\mu . \tag{10.6}$$

Proof. (a) *The condition is sufficient.* Let $g \in Y$. Since $f - g_0$ is 0 on A_0, it follows from (10.6) in the L_1 norm,

$$\begin{aligned} \|f - g_0 - g\| &\ge \int_{A \backslash A_0} (f - g_0 - g) \operatorname{sign}(f - g_0) d\mu + \int_{A_0} |g| d\mu \\ &= \|f - g_0\| + \int_{A_0} |g| d\mu - \int_A g \operatorname{sign}(f - g_0) d\mu \\ &\ge \|f - g_0\| . \end{aligned}$$

(b) *The condition is necessary.* Let $g_0 \in Y$ be a best approximation to f and let h be the function of (ii) in Corollary 10.2. Then $h = \operatorname{sign}(f - g_0)$ on $A \backslash A_0$ and $\|h\|_\infty = 1$. From the orthogonality of h, we have for any $g \in Y$,

$$\int_A g \operatorname{sign}(f - g_0) d\mu = -\int_{A_0} g h d\mu \le \int_{A_0} |g| d\mu . \qquad \square$$

For the case $A := [a, b]$, we shall now describe some special functions h that are orthogonal to Y and can be used to explicitly determine best L_1 approximants for certain functions $f \in L_1$. Let $(a_k)_{k=1}^n$ form a partition of $A : a =: a_0 < a_1 < \cdots < a_n < a_{n+1} := b$. The sequence (a_k) is called *canonical* if the signature

$$\sigma(x) := (-1)^k , \quad a_k < x < a_{k+1} , \quad k = 0, \ldots, n \tag{10.7}$$

is orthogonal to Y : $\int_A g \sigma d\mu = 0$ for all $g \in Y$. A long time ago, Markov noticed the importance of canonical points and described them in some cases.

Theorem 10.5. *Let f be continuous and let (a_k) be a sequence of canonical points for Y. If a continuous function $g_0 \in Y$ interpolates f at the points a_k, $k = 1, \ldots, n-1$, and if $f - g_0$ changes sign at these points and no others, then g_0 is a best approximation to f in L_1 and the approximation error is*

$$E(f, y)_{L_1} = \left| \int_A f \sigma d\mu \right| . \tag{10.8}$$

Proof. One of the functions $h = \sigma$ or $h = -\sigma$ satisfies (ii) of Corollary 10.2. $\square$

The existence of canonical points follows from the theorem of Hobby and Rice [1965] below. We shall use:

Theorem (Borsuk) (for a proof see DeVore, Kierstead, and Lorentz [1987]). *Let S^n be a sphere in the space $\mathbb{R}^{n+1}$ equipped with some norm and let T be a continuous mapping of S^n into $\mathbb{R}^n$ that is odd, $T(-y) = -T(y)$. Then $T(y^*) = 0$ for some $y^* \in S^n$.*

Theorem 10.6. *Let μ be a non-atomic measure on $A = [a, b]$. For each n-dimensional subspace $Y \subset L_1(A)$ there is a sequence of $m \le n$ points that are canonical for Y.*

Proof. Let S^n be the Euclidean sphere of points $x = (x_0, \ldots, x_n)$ satisfying $\sum_0^n x_k^2 = b - a$. We define $a(x) := (a_k(x))_0^{n+1}$ by means of $a_0(x) := a$, $a_1(x) := a + x_0^2, \ldots, a_{n+1}(x) := a + x_0^2 + \cdots + x_n^2$. If Y is spanned by a basis $g_1, \ldots, g_n$, we define the mapping $T(x) := (T_i(x))_1^n$ of S^n into $\mathbb{R}^n$ by means of

$$T_i(x) := \sum_{k=0}^n \operatorname{sign} x_k \int_{a_k(x)}^{a_{k+1}(x)} g_i d\mu , \quad i = 1, \ldots, n . \tag{10.9}$$

The mapping T is clearly odd and since μ is non-atomic and a is continuous, T is also continuous. Hence from Borsuk's theorem there is a point x^* with $T_i(x^*) = 0$, $i = 1, \ldots, n$. We throw out from (10.9) all terms with $x_k^* = 0$ (they are over degenerate intervals) and combine adjacent terms with equal sign x_k^*. In this way, we obtain a signature σ with $m \le n$ breakpoints, orthogonal to all the h_i. $\square$

With Theorem 10.6, we can prove the following remarkable theorem of Krein [1938].

Theorem 10.7. *Let μ be a non-atomic measure on $A = [a, b]$. No finite dimensional subspace Y of $L_1(A)$ is a unicity space for $L_1(A)$.*

Proof. Let (a_k) be a sequence of canonical points for Y and $\sigma(x)$ be its signature (10.7). We take a function $g \in Y$ which is not identically zero and set $f := \sigma|g|$. By Corollary 10.2 (ii), 0 is a best approximation to f. Moreover, for any $\alpha \le 1$, we have $|\alpha\sigma| \le 1$ and

$$\sigma(f - \alpha g) = \sigma(\sigma|g| - \alpha g) = |g| - \alpha\sigma g \ge 0 .$$

Therefore, αg is also a best approximation to f from Y. $\square$

For the remainder of this section, we shall consider the case where Y has a basis which is a Haar system $\Phi := \{\phi_1, \ldots, \phi_n\}$ on $A = [a, b]$ or $\mathbb{T}$. The theory here is much richer. We shall restrict ourselves to measures $d\mu = w(x)dx$ with $w(x) > 0$, $x \in A$.

If $f \in L_p$, $p > 1$, the functional λ_f of (10.1) has the representation

$$\lambda_f(g) := \int_A gh d\mu \text{ where } \operatorname{sign} h = \operatorname{sign} f, \text{ a.e.} \tag{10.10}$$

(with h from (10.3)). This is also true for $p = 1$ for one of the functionals λ_f.

The following theorem shows that the error of approximation must oscillate sufficiently often.

Theorem 10.8. (i) *Let $f \in L_p(A)$, $1 < p < \infty$, be continuous and let $P \neq f$ be a Φ-polynomial of best approximation to f. Then, $f - P$ changes sign at least n times on A if $A = [a, b]$, at least $n + 1$ times on A if $A = \mathbb{T}$.* (ii) *This remains true also for the space L_1, if the set where $f(x) = P(x)$ has measure zero.*

Proof. (i) We treat only the case $A = [a, b]$; the case $A = \mathbb{T}$ is similar. We suppose that $f - P$ has only $k < n$ sign changes in A and we derive a contradiction. There are points $a =: x_0 < x_1 < \cdots < x_{k+1} := b$ and an $\varepsilon = \pm 1$ such that $\varepsilon(-1)^i(f(x) - P(x)) \geq 0$, $x \in (x_i, x_{i+1})$, $i = 0, \ldots, k$. By Theorem 9.1, there is a polynomial Q which changes sign exactly at the points $x_1, \ldots, x_k$ so that $(f - P)$ and Q are never of opposite sign. Let $h^* := h_{f-P}$ be the extremal function of Corollary 10.2 with $\operatorname{sign} h^* = \operatorname{sign}(f - P)$. Then, Q and h^* are never of opposite sign and there are subintervals where $Q(x)h^*(x) > 0$. Therefore $\int Qh^* d\mu > 0$ which is a contradiction to Corollary 10.2 (i). The proof of (ii) follows similarly from Corollary 10.2 (ii). □

In this situation, we have as a corollary that a Φ-polynomial of best approximation to f *interpolates* f at at least n interior points of $[a, b]$ or at at least $n + 1$ points of $\mathbb{T}$. We also obtain the following uniqueness theorem for L_1 approximation.

Theorem 10.9 (Jackson [A-1930]). *In the space $L_1(A)$, $A = [a, b]$ or $A = T$, each continuous function has a unique Φ-polynomial of best approximation.*

Proof. Suppose that P_1 and P_2 are both polynomials of best L_1 approximation to the continuous function f. From our remark (1.4), $(P_1 + P_2)/2$ is also a best approximation and

$$|2f(x) - P_1(x) - P_2(x)| = |f(x) - P_1(x)| + |f(x) - P_2(x)|, \text{ a.e.} \tag{10.11}$$

Since the functions under consideration are continuous, (10.11) holds for all x. From Theorem 10.8, the left side of (10.11) vanishes at at least n points. Hence, P_1 and P_2 agree at these points and therefore $P_1 = P_2$. □

Assume now that both $\Phi = \{\phi_1, \ldots, \phi_n\}$ and $\Phi' = \{\phi_1, \ldots, \phi_n, \phi_{n+1}\}$ are Haar systems on $A = [a, b]$. Let $\widetilde{P}_n$ be the Φ-polynomial of best L_1 approximation to the function ϕ_{n+1}. Since $\phi^*_{n+1} := \phi_{n+1} - \widetilde{P}_n$ is a nontrivial Φ'-polynomial, it follows from Theorem 10.8 that ϕ^*_{n+1} has on A precisely n zeros $x_1 < \cdots < x_n$ where it changes sign. Then from Corollary 10.2 (ii), *the points x_k are canonical for the space Y spanned by Φ and the corresponding signature is $\sigma = \operatorname{sign} \phi^*_{n+1}$.* Therefore, Theorem 10.5 applies. In particular, the assumptions of the theorem are satisfied if $\phi_1, \ldots, \phi_n, f$ is a Haar system.

It is also noteworthy in this situation that the canonical sequence is unique for Y. Indeed, if (x'_k) were a second canonical sequence, then, by Theorem 10.5, the polynomial Q which interpolates ϕ_{n+1} at the x'_k, $k = 1, \ldots, n$, is also a best approximation to ϕ_{n+1}. By Theorem 10.8, best approximation to ϕ_{n+1} is unique and therefore $Q = \widetilde{P}_n$ and $x_k = x'_k$, $k = 1, \ldots, n$.

As an example, we consider approximation by the elements of the space $\mathcal{P}_{n-1}$ of algebraic polynomials of degree $< n$. We can describe the canonical sequence with the aid of the Chebyshev polynomials C_n of §6. The polynomials $U_n := C'_{n+1}/(n + 1)$, given also by

$$U_n(x) = \frac{\sin(n+1)t}{\sin t}, \quad x = \cos t, \quad -1 \leq x \leq 1 \tag{10.12}$$

are called the *Chebyshev polynomials of the second kind.* The leading coefficient of U_n is 2^n. A change of variables $x = \cos t$ and Lemma 10.3 yield for each P of degree $\leq n - 1$,

$$\begin{aligned}\int_{-1}^{+1} P(x) \operatorname{sign} U_n(x) dx &= \int_0^\pi P(\cos t) \operatorname{sign}\left[\frac{\sin(n+1)t}{\sin t}\right] \sin t\, dt \\ &= \int_0^\pi P(\cos t) \sin t \ \operatorname{sign}[\sin(n+1)t] dt \\ &= \frac{1}{2}\int_{-\pi}^{\pi} P(\cos t) \sin t \ \operatorname{sign}[\sin(n+1)t] dt = 0\,.\end{aligned}$$

Therefore, the zeros of U_n

$$x_k = \cos\frac{k\pi}{n+1}, \quad k = 1, \ldots, n, \tag{10.13}$$

are a canonical sequence for $\mathcal{P}_{n-1}$ in $[-1, 1]$.

Theorem 10.10 (Markov). *If $f \in C^{n-1}[-1, 1]$ and $f^{(n)}$ exists and does not vanish on $(-1, +1)$, then the algebraic polynomial of best L_1 approximation to f of degree $< n$ is the polynomial which interpolates f at the points* (10.13) *and*

$$E_n(f)_1 = \left|\int_{-1}^{1} f \operatorname{sign}\ U_n dx\right|. \tag{10.14}$$

Proof. This follows from the previous discussion because $1, x, \dots, x^{n-1}, f$ is a Haar system. □

In particular, taking $f(x) = x^n$, we deduce that $2^{-n}U_n$ is the polynomial of degree n with leading coefficient 1 of least deviation from zero on $[-1,1]$ in the L_1 norm. Since $\int_{-1}^{1} |U_n| dx = \frac{1}{n+1} \operatorname{Var} C_{n+1} = 2$, we get also

$$E_{n-1}(x^n)_{L_1} = 2^{-n+1} \text{ on } [-1,1]. \tag{10.15}$$

§ 11. Pólya and Descartes Systems

Special cases of Haar systems on $[a,b]$ are the Pólya systems. Their definition is based on the following remarks:

1. If $V = \{v_0, \dots, v_{k-1}\}$ is a Haar system on $[a,b]$ and $w \in C[a,b]$ is strictly positive, then $W := \{wv_0, \dots, wv_{k-1}\}$ is also a Haar system on $[a,b]$ and the determinants $D(V; x_0, \dots, x_{k-1})$ and $D(W; x_0, \dots, x_{k-1})$ of (3.3) have the same sign.

2. For each c, with $a \le c \le b$, the system U given by

$$u_0(x) := 1, \quad u_j(x) := \int_c^x v_{j-1}(t)\, dt, \quad j = 1, \dots, k \tag{11.1}$$

is a Haar system on $[a,b]$, and the determinants $D(V; t_1, \dots, t_k)$ and $D(U; x_0, \dots, x_k)$ for $a \le t_1 < \cdots < t_k \le b$, $a \le x_0 < \cdots < x_k \le b$ have the same sign.

For the proof of **2**, we write the second determinant in the form (3.3). We replace $u_j(x_i)$ in the i-th row by $\int_c^{x_i} v_{j-1}(t_j) dt_j$, for $j = 1, \dots, k$, by 1 for $j = 0$. We then subtract row i from row $i+1$ for $i = k-1, \dots, 0$. We obtain a $k \times k$ determinant, and taking the integrals outside we get

$$D(U; x_0, \dots, x_k) = \int_{x_0}^{x_1} \cdots \int_{x_{k-1}}^{x_k} D(V; t_1, \dots, t_k)\, dt_1 \dots dt_k. \tag{11.2}$$

This proves the assertion because $x_0 < \cdots < x_k$ implies $t_1 < \cdots < t_k$ inside the domain of integration.

A *Pólya system* V is defined for strictly positive functions $w_0, \dots, w_n \in C[a,b]$ by the formulas

$$\begin{cases} u_0(x) = w_0(x) \\ u_1(x) = w_0(x) \displaystyle\int_c^x w_1(t_1) dt \\ \dots\dots\dots\dots\dots\dots \\ u_n(x) = w_0(x) \displaystyle\int_c^x w_1(t_1) \int_c^{t_1} w_2(t_2) \cdots \int_c^{t_{n-1}} w_n(t_n) dt_n \dots dt_1 \end{cases} \tag{11.3}$$

where c, with $a \le c \le b$, is fixed. That this is a Haar system, follows from **1** and **2**. We start with the system consisting of one function 1, and applying the operations **1, 2** several times, we end up with the system (11.3). We have

3. For each Pólya system V, one has

$$D(V; x_0, \dots, x_n) > 0 \text{ when } a \le x_0 < \cdots < x_n \le b.$$

This follows by induction from **1** and **2.**

The simplest example of (11.3) is when $w_k(x) \equiv 1$, $k = 0, \dots, n$. Then the system becomes $1, (x-c), \dots, \frac{1}{n!}(x-c)^n$.

Our definitions also apply to Pólya systems on an open interval (a,b).

There is another way of obtaining the systems (11.3), which has been Pólya's point of departure. Let $w_k(x) > 0$, be $(n-k)$ times continuously differentiable on $[a,b]$. We define the linear differential operators D_k and D by

$$\begin{cases} D_k g = \left(\dfrac{1}{w_k} g\right)', & k = 0, \dots, n-1, \\ Dg = \dfrac{1}{w_n} D_{n-1} \dots D_0 g. \end{cases} \tag{11.4}$$

They act on functions $g \in C^n[a,b]$. One sees that the functions $u_0, \dots, u_{n-1}$ of (11.3) satisfy $Du = 0$. Since they are linearly independent, they span the null space of D. On the other hand, u_n is a solution of $Du = 1$.

Another special case are the Descartes systems introduced (with a different but equivalent definition) by Bernstein ([A-1926, p. 27]). A set $\Phi := \{\phi_1, \dots, \phi_n\}$ of continuous functions is a *Descartes system* on $[a,b]$, if for each $k = 1, \dots, n$, there is a sign $\varepsilon_k = \pm 1$, so that the determinants

$$D(\lambda_1, \dots, \lambda_k;\ x_1, \dots, x_k) = \begin{vmatrix} \phi_{\lambda_1}(x_1) & \dots & \phi_{\lambda_k}(x_1) \\ \dots & \dots & \dots \\ \phi_{\lambda_1}(x_k) & & \phi_{\lambda_k}(x_k) \end{vmatrix} \tag{11.5}$$

with $1 \le \lambda_1 < \cdots < \lambda_k \le n$, $a < x_1 < \cdots < x_k < b$ have all the sign ε_k. Thus, each set $\{\phi_{\lambda_1}, \dots, \phi_{\lambda_k}\}$ is a Haar system on (a,b).

The main example is:

$$\begin{cases} \textit{The functions } 1, x, \dots, x^{n-1} \textit{ are a Descartes system on } [a,b] \\ \textit{for each } a > 0 \textit{ with values } \varepsilon_k = 1,\ k = 1, \dots, n. \end{cases} \tag{11.6}$$

Indeed, from Example 2 of §3, all determinants (11.5) are $\ne 0$. Since they are all continuous in the real variables λ_i, the sign of the determinant (11.5) is that of the Vandermonde determinant $D(0, \dots, k-1;\ x_1, \dots, x_k) > 0$.

It follows from the definition that with $\{\phi_1, \dots, \phi_n\}$ also the set $\{\psi_1, \dots, \psi_k\}$ is a Descartes system, if $\psi_j := A_{m_j}\phi_{m_j} + \cdots + A_{m_{j+1}-1}\phi_{m_{j+1}-1}$, $j = 1, \dots, k$, with $1 = m_1 < \cdots < m_{k+1} = n+1$ and positive coefficients A_i.

We shall study the non-trivial polynomials

$$P(x) = \sum_{i=1}^{k} a_i \phi_{\lambda_i}(x), \quad 1 \le \lambda_1 < \cdots < \lambda_k \le n, \tag{11.7}$$

which interpolate zero at some fixed points $a < x_1 < \cdots < x_{k-1} < b$:

$$P(x_i) = 0, \qquad i = 1, \ldots, k-1. \tag{11.8}$$

It follows from §3 that such P exist; they are defined up to a constant factor. They have no zeros different from the x_i, and change sign at these points.

Lemma 11.1. *The coefficients a_i of the polynomials* (11.7) *are different from zero and satisfy*

$$\operatorname{sign}(a_i a_{i+1}) = -1, \qquad i = 1, \ldots, k-1 \tag{11.9}$$

$$\varepsilon_{k-1}\varepsilon_k a_k P(x) > 0, \quad \textit{for } x_{k-1} < x < b. \tag{11.10}$$

Proof. Let $x_{k-1} < x < b$. To the equations (11.8), we add the equation (11.7) and solve for the a_i. By Cramer's rule, we obtain for a_i the quotient of two determinants with signs ε_{k-1} and ε_k multiplied by $(-1)^{i+k}P(x)$. With the same x, for a_{i+1}, we obtain an expression of opposite sign. This yields (11.9). For $i = k$, we obtain (11.10). □

From this, we derive a comparison theorem of Pinkus and Smith (see Smith [1978]).

Theorem 11.2. *Let the polynomials P, Q interpolate zero at the same $k-1 < n$ points:*

$$P(x_i) = Q(x_i) = 0, \quad i = 1, \ldots, k-1; \quad a < x_1 < \cdots < x_{k-1} < b$$

and have the form

$$\begin{cases} P(x) = \sum_{i=1}^{k-1} a_i \phi_{\lambda_i} + \phi_n, & 1 \le \lambda_1 < \cdots < \lambda_{k-1} < n \\ Q(x) = \sum_{i=1}^{k-1} b_i \phi_{\mu_i} + \phi_n, & 1 \le \mu_1 < \cdots < \mu_{k-1} < n. \end{cases} \tag{11.11}$$

If $P \ne Q$ and $\mu_i \le \lambda_i$, $i = 1, \ldots, k-1$, then

$$|P(x)| < |Q(x)|, \quad x \in (a,b), \quad x \ne x_i, \quad i = 1, \ldots, k-1. \tag{11.12}$$

Proof. The polynomials P and Q exist and are uniquely defined. It will be sufficient to prove (11.12) for the case when there is only one index $i = i_0$ where μ_i, λ_i are different, so that $\lambda_{i_0-1} < \mu_{i_0} < \lambda_{i_0}$. The general case is reduced to this by repetition. Then P, Q, $P - Q$ are of the form

$$P = \sum_i a_i \phi_{\lambda_i} + \phi_n, \qquad Q = \sum_i b_i \phi_{\mu_i} + \phi_n$$

$$P - Q = \sum_{i \ne i_0} c_i \phi_{\lambda_i} - b_{i_0}\phi_{\mu_{i_0}} + a_{i_0}\phi_{\lambda_{i_0}}$$

and all coefficients of $P - Q$ are non-zero because $P - Q$ has $k-1$ zeros. By (11.9) applied to P, $\operatorname{sign}(1\, a_{i_0}) = (-1)^{k-i_0}$ and by (11.9) for $P - Q$, if c is the leading coefficient of $P - Q$ (which is c_{k-1} if $i_0 \ne k-1$), $\operatorname{sign}(c\, a_{i_0}) = (-1)^{k-1-i_0}$. Hence, $c < 0$. If we write $\varepsilon := \varepsilon_{k-1}\varepsilon_k$, then from (11.10) applied to P, Q and $P - Q$, we obtain for $x_{k-1} < x < b$:

$$\varepsilon P(x) > 0, \quad \varepsilon Q(x) > 0, \quad \varepsilon\,(P(x) - Q(x)) < 0.$$

Therefore, $0 < \varepsilon P(x) < \varepsilon Q(x)$ on this interval, that is, we have (11.12). At $x = x_{k-1}$, all polynomials change sign and for $x_{k-2} < x < x_{k-1}$, we get

$$\varepsilon P(x) < 0, \quad \varepsilon Q(x) < 0, \quad \varepsilon\,(P(x) - Q(x)) > 0,$$

which is again (11.12). Continuing in this way, we can extend this inequality to the whole of (a,b). □

§12. Weak Haar Systems

The definition of a Haar space given in §3 is in several cases too restrictive. For example, the Schoenberg spline spaces of Chapter 5 do not quite fit into this category. However, several important properties of Haar systems remain preserved in a somewhat weaker form, for more general systems. Jones and Karlovitz [1970] proved Theorem 12.1 below for continuous functions on $[a,b]$. This theorem remains valid for discontinuous functions defined on an arbitrary subset A of $\mathbb{R}$ that has at least $n+1$ points.

A set $\Phi : \phi_1, \ldots, \phi_n$ of real valued functions on A is a *weak Haar system* on A if the functions ϕ_j are linearly independent and if there does not exist a non-trivial polynomial $P = \sum_1^n a_j \phi_j$ which has an alternating sign sequence of length $n+1$, that is a sequence $X : x_0 < \cdots < x_n$, $X \subset A$, with

$$\sigma(-1)^i P(x_i) > 0, \quad i = 0, \ldots, n, \text{ where } \sigma = 1 \text{ or } \sigma = -1 \tag{12.1}$$

The space $\mathcal{F}$ spanned by Φ is a *weak Haar space* on A.

For example, a Haar system Φ of continuous functions on $A = [a,b]$ has this property, for (12.1) implies that P has n distinct zeros on A. The converse is not true. The trigonometric system $1, \cos x, \ldots, \sin nx$ is a weak Haar system on the interval $[0, 2\pi]$ (without identification of the endpoints) but is not a Haar system, for $\sin nx$ has $2n+1$ zeros on $[0, 2\pi]$. An equivalent condition can be given in terms of the constancy of the sign of the determinants for $x_j \in A$, $j = 1, \ldots, n$,

$$D(X) := D(\Phi; X) := D(\Phi; x_1, \ldots, x_n) := \det\,(\phi_j(x_i)), \quad X : x_1 < \cdots < x_n. \tag{12.2}$$

If $D(X) \neq 0$ for some n-tuple X, then we can interpolate arbitrary data by means of the values of $P \in \mathcal{F}$ at the x_i. In particular, there exist Lagrange polynomials L_j, $j = 1, \dots, n$,

$$L_j(x_i) = \delta_{i,j}\,, \quad i, j = 1, \dots, n\,.$$

The L_j are another basis for $\mathcal{F}$.

The following thoerem is due to Jones and Karlovitz [1970] for the case of continuous functions ϕ_j. In this case, one can approximate $\mathcal{F}$ by ordinary Haar spaces (see §7 of Chapter 10). The proof below (by Zielke [A-1979]) uses only tools of linear algebra.

Theorem 12.1. *A system Φ of n linearly independent real function on $A \subset \mathbb{R}$ (A contains at least $n+1$ points) is a weak Haar system if and only if no two determinants* (12.2) *are of opposite sign.*

Proof. This will follow at once form the statements (i) and (ii) that follow.

(i) *If $P \in \mathcal{F}$, $P \neq 0$, has an alternation of signs of length $n+1$:*

$$(-1)^i P(x_i) > 0\,, \quad X:\ x_0 < \cdots < x_n\,, \tag{12.3}$$

then some two determinants (12.2) *are of opposite sign.*

First of all, we reduce the problem to the case when the space $\mathcal{F}$ restricted to X has dimension n. Indeed, if this dimension is equal to $k < n$ we shall construct a modified $n+1$- tuple X^* and a polynomial $P^* \in \mathcal{F}$ with full sign alternation on X^* such that $\mathcal{F}$ restricted to X^* is of dimension $k+1$.

Since $\mathcal{F}$ has dimension n, there is a $y \in A$, different from each $x_i \in X$ so that $\mathcal{F}$ has dimension $k+1$ on $X \cup \{y\}$. We choose μ so that $x_\mu < y < x_{\mu+1}$. There are indices ν such that $\mathcal{F}$ has dimension k on $X \setminus \{x_\nu\}$. We choose such a point with ν closest to μ. We assume that $\nu > \mu$; the other case is proved similarly.

The new $n+1$-tuple X^* is obtained by dropping x_ν from X and adding y, so that $x^*_{\mu+1} := y$, $x^*_j := x_{j-1}$, for $\mu + 1 < j \le \nu$, and $x^*_j := x_j$ for all other j. Then $\mathcal{F}$ has dimension $k+1$ when restricted to X^*. Also by our definition of ν, $\mathcal{F}$ has dimension $\nu - \mu + 1$ when restricted to $\{x^*_\mu, \dots, x^*_\nu\}$.

To the points in $\{x^*_\mu, \dots, x^*_\nu\}$, we add $k + 1 - \nu + \mu - 1$ additional points from X^* so that the resulting set Y has $k+1$ points and $\mathcal{F}$ has dimension $k+1$ when restricted to Y. We can then adjoin to Y, $n - k + 1$ additional points from A to obtain a set Z such that $\mathcal{F}$ has dimension n when restricted to Z. We take the Lagrange basis (with respect to Z) for $\mathcal{F}$. Let L_z denote the Lagrange function corresponding to $z \in Z$. We claim that for $\mu \le j \le \nu$

$$L_{x^*_j}(x^*_i) = 0\,, \quad i = 0, \dots, n\,, \quad i \neq j\,. \tag{12.4}$$

Consider first the case $j \neq \mu+1$. Then, (12.4) is clear if $x^*_i \in Y$. On the other hand, if $x^*_i \notin Y$, and $L_{x^*_j}(x^*_i) \neq 0$, then the functions L_z, $z \in Z$, are linearly independent over $(Z \cup \{x^*_i\}) \setminus \{x^*_j\}$. But this implies that $\mathcal{F}$ has dimension $k+1$ when restricted to $X^* \setminus \{x^*_j\}$ which contradicts the definition of ν. The same argument, using now the definition of y, shows that (12.4) is valid for $j = \mu + 1$. Thus (12.4) has been verified.

Now let

$$Q(x) := \sum_{j=\mu}^{\nu} (-1)^j L_{x^*_j}(x)\,.$$

Then, Q takes the values $(-1)^j$ at the points x^*_j, $\mu \le j \le \nu$, and takes the value 0 at all other x^*_j. Therefore $P^* := Q + \varepsilon P$, with P satisfying (12.3), has full sign alternation on X^* provided $\varepsilon > 0$ is sufficiently small.

Repeating the above argument, we arrive at an $n+1$-tuple X on which $\mathcal{F}$ has dimension n when restricted to X and a function $P \in \mathcal{F}$ which has full alternation on X. Eliminating the c_j from the equations

$$\sum_{j=1}^{n} c_j \phi_j(x_i) = P(x_i)\,, \quad i = 0, \dots, n+1\,,$$

we obtain

$$\sum_{i=0}^{n} (-1)^i P(x_i) D_i = 0\,, \quad D_i := D(\Phi; x_0, \dots, x_{i-1}, x_{i+1}, \dots, x_n)\,. \tag{12.5}$$

One of the D_i is not zero. Since all the coefficients in (12.5) satisfy $(-1)^i P(x_i) > 0$, some two of the D_i are of opposite signs and (i) is proved.

(ii) *If two determinants* (12.2) *for the n-tuples $X : x_1 < \cdots < x_n$ and $Y : y_1 < \cdots < y_n$ are of opposite sign:*

$$D_1 D_2 < 0\,, \quad D_1 := D(\Phi; X)\,, \quad D_2 := D(\Phi; Y)\,, \tag{12.6}$$

then some $P \in \mathcal{F}$ has n sign alternations on A.

We note that property (12.6) is invariant under a change of basis for the space $\mathcal{F}$. Indeed, this change will convert D_i into ΔD_i, $i = 1, 2$ for some real number $\Delta \neq 0$.

We can obtain D_2 from D_1 by at most n operations of the following type. At each step, we drop a row of D_1 corresponding to an x_p not equal to one of the y_i and replace it by the row $(\phi_j(y))_{j=1}^n$ (placed in its proper position), where $y := y_q$ is selected so that this row is linearly independent of the remaining $n-1$ rows of the current D_1. The determinants obtained in this way will be all different from zero. Consequently, some two of them will be of opposite sign.

We have proved that there exists a pair D_1, D_2 of (12.6) where the y_i are the same as the x_i except that x_p is missing and exactly one $y := y_q$ is none of the x_i. Now, we replace the ϕ_j by the Lagrange functions L_j corresponding to the points X. Then $D_1 = 1$, and therefore $D_2 = \det(L_j(y_i)) < 0$. Thus, L_p has the properties

$$L_p(x_p) = 1\,, \quad L_p(x_i) = 0\,, \quad i \neq p\,, \quad L_p(y) = c(-1)^\sigma\,, \tag{12.7}$$

where $c > 0$ and $\sigma := |p - q| + 1$.

Let $z_1 < \cdots < z_{n+1}$ be the $n+1$ points x_i and y in ascending order. Assume that $p < q$, then $y = z_{q+1}$. We define signs $\varepsilon_j = \pm 1$ by putting $\varepsilon_p := 1$, $\varepsilon_q := (-1)^\sigma$, and then extending ε_j to all $j = 1, \ldots, n+1$ to have a fully alternating sequence (this also works if $p \geq q$).

Since $D_2 \neq 0$, there is a $Q \in \mathcal{F}$ for which

$$Q(z_j) = \varepsilon_j, \quad \text{for all } j, \quad j \neq p.$$

Then for sufficiently small $\varepsilon > 0$, $L_p + \varepsilon Q$ will have the signs ε_j at z_j, $j = 1, \ldots, n+1$. This is the required $P \in \mathcal{F}$ of (ii). □

For a Haar space $\mathcal{F}$ of dimension n, polynomials $P \in \mathcal{F}$, $P \neq 0$, can have at most $n-1$ zeros. For a weak Haar space, the number of zeros may be infinite. Indeed, the extension by zero onto $[b, c]$ of all functions of a weak Haar space on $[a, b]$ produces another weak Haar space. We have to change the way of counting zeros to obtain a sensible upper bound for their number.

A point $x \in A$ is an *essential point* of a weak Haar system $\Phi: \phi_1, \ldots, \phi_n$ on A if not all of the ϕ_j vanish at x. A zero is essential if it is an essential point. Two zeros $x_1 < x_2$ of $P \in \mathcal{F}$ are separated, if for some $y \in A$, $x_1 < y < x_2$, $P(y) \neq 0$. With these definitions, we have the following theorem (proved by Bartlet [1975] for continuous functions on an interval):

Theorem 12.2. (i) *An element P of an n-dimensional weak Haar space $\mathcal{F}$ cannot have more than n essential separated zeros.* (ii) *If $x_1 < \ldots x_n$ are n essential separated zeros of P then $P(x) = 0$ for all $x \in A$, $x < x_1$, and for all $x \in A$, $x > x_n$.*

Proof (by Schumaker). It is sufficient to prove (ii). We first prove by induction, that if $x_1 < \cdots < x_m$ is a finite number of essential points of A, then there is an element $P \in \mathcal{F}$ for which $P(x_i) \neq 0$, $i = 1, \ldots, m$. If $m = 1$, we can take for P one of the ϕ_i which does not vanish at x_1. Suppose that our statement is true for $m-1$ and P has been constructed for $x_1 < \cdots < x_{m-1}$. If $P(x_m) \neq 0$, we have the desired polynomial. On the other hand, if $P(x_m) = 0$, we take ϕ_j with $\phi_j(x_m) \neq 0$. Then for sufficiently small $\varepsilon \neq 0$, $P + \varepsilon\phi_j$ is the desired polynomial.

Now assume that (ii) does not hold; for instance that P does not vanish identically to the right of x_n. Then, there are points

$$x_1 < y_1 < \cdots < y_{n-1} < x_n < y_n \tag{12.8}$$

so that the x_i are essential zeros of P and $P(y_i) \neq 0$, $i = 1, \ldots, n$. We take a function $Q \in \mathcal{F}$ for which $Q(x_i) \neq 0$, $i = 1, \ldots, n$. The two sequences

$$Q(x_1), P(y_1), \ldots, P(y_{n-1}), Q(x_n), P(y_n), \tag{12.9}$$

$$Q(x_1), -P(y_1), \ldots, -P(y_{n-1}), Q(x_n), -P(y_n) \tag{12.10}$$

have altogether exactly $2n-1$ changes of sign. One of them – let it be the first sequence – has at least n sign changes. If $z_0 < z_1 < \cdots < z_n$ is a subsequence of (12.9) which realizes n changes of sign, then for small $\varepsilon \neq 0$, $P^* := P + \varepsilon Q$ alternates in sign at the z_j, in contradiction to the definition of a weak Haar space. □

For further information about weak Haar systems see Chapter 5, §8 and Chapter 10, §8.

§ 13. Problems

13.1. Functions $f(x, y) \in C^1(D)$, where D is the closed unit disk in $\mathbb{R}^2$, have unique best uniform approximations by linear functions $Ax + By + C$ (Collatz).

13.2. The functions $x^{k_0}, \ldots, x^{k_n}$, where $0 \leq k_0 < \cdots < k_n$ are integers, are a Haar system on $[-1, 1]$ if and only if $k_0 = 0$ and all differences $k_i - k_{i-1}$ are odd (Lorentz, Passow, see the book of Lorentz, Jetter, and Riemenschneider [A-1983, p. 132].

13.3. The functions $\frac{1}{x+c_1}, \ldots, \frac{1}{x+c_n}$, $c_1 < \cdots < c_n$ are a Haar system on $[a, b]$ provided $a + c_1 > 0$.

13.4. The best uniform approximation P by linear functions to a convex function f on $[-1, 1]$ is $P = l - c/2$ where l is the linear interpolant to f at ± 1 and $c := \|f - l\|$.

13.5. Among all polynomials of degree $\leq n$, zero is the best uniform approximation to z^{n+1} on $|z| \leq 1$.

13.6. The polynomial of degree $2n - 2$ which interpolates $f(x) = |x|$ at the zeros of U_{2n} is the best L_1-approximation to f on $[-1, 1]$ among all polynomials of degree $\leq 2n - 1$. (DeVore [1968]).

13.7. If $\Phi : \phi_1, \ldots, \phi_n \in C[a, b]$ are linearly independent, then there exist points $X : x_1, \ldots, x_n$ for which Φ is a Haar system on X.

13.8. The coefficients of $C_n(x) = \sum_{k=0}^{n} a_k x^k$ satisfy the recurrence relation $a_k(n^2 - k^2) + a_{k+2}(k+1)(k+2) = 0$, $k = 0, \ldots, n-2$.

13.9. (Wulbert [1971]) If for $f \in C(A)$, $A = [a, b]$, a function P from the n-dimensional subspace X_n of $C(A)$ satisfies the Kolmogorov criteria (2.4) with strict inequality, then P is the strongly unique best approximation to f.

§ 14. Notes

14.1. The theorem of Helly for convex sets (see Eggleston [B-1958, p. 33]) is sometimes useful in approximation. This theorem says that a collection $\mathcal{K}$ of compact convex sets $K \subset \mathbb{R}^n$ has a non-empty intersection provided every subcollection consisting of $n+1$ sets from $\mathcal{K}$ has this property. This is quite simple in the case $n = 1$ since then the sets are necessarily intervals. For $n > 1$, it can be proved by induction.

14.2. A corollary to Theorem 5.1 is that the error of approximation of f by a Haar space of dimension n on $[a, b]$ is equal to its error of approximation on some properly chosen $n+1$ points of the interval. This fact is valid in a much more general setting.

Theorem (Shnirelman [1938]). *Let X_n be an n-dimensional subspace of $C(A)$, A compact Hausdorff. If $f \in C(A)$, there is a subset B_0 of A, consisting of at most $n+1$ points, so that $\varepsilon := E_n(f, A) = E_n(f, B_0)$.*

Proof. Let f be fixed. The subsets $B = \{x_1, \ldots, x_{n+1}\}$ of A with $n+1$ points (with possible repetitions) can be interpreted as points in the product space A^{n+1}.

Let $\varepsilon' := \sup_B E_n(f, B)$. A compactness argument shows that this supremum is attained for some $n+1$ tuple B_0.

For any $x \in A$, let $\Omega_x := \{P \in X_n : |f(x) - P(x)| \le \varepsilon'\}$; this is a closed and convex subset of X_n. We can identify Ω_x with a subset of $\mathbb{R}^n$ by taking a basis for X_n and identifying $P \in X_n$ with its coefficients in its representaiton with respect to that basis. By the definition of ε', for any $n+1$ tuple $\{x_1, \dots, x_{n+1}\}$, there exists a $P \in \Omega_{x_1} \cap \dots \cap \Omega_{x_{n+1}}$. We use Helly's theorem and obtain the existence of a P which is in each of the Ω_x, $x \in A$. Hence, $|f(x) - P(x)| \le \varepsilon'$, $x \in A$, so that $\varepsilon \le \varepsilon'$. Since the inequality $\varepsilon' \le \varepsilon$ is trivial, we get $\varepsilon = \varepsilon'$. □

14.3. We have found the polynomial P_{n-1} of best uniform approximation to x^n by *guessing* it. Chebyshev [B-1962] gives an elegant direct approach. For the best P_{n-1} and $y := x^n - P_{n-1}(x)$, by Theorem 5.1, $|y|$ has at least $n+1$ maxima on $[-1,1]$ with the same value d_n. Since y' has at most $n-1$ zeros, there are exactly $n+1$ maxima and two of them are at ± 1. It follows that

$$d_n^2 - y^2 = \gamma_n (1 - x^2) y'(x)^2$$

since the right and left hand sides have the same zeros. Comparing the coefficients, we see that $\gamma_n = n^{-2}$. Hence, from this differential equation we can find y, P_{n-1} and d_n.

This method can be useful in other cases. In this way, Bernstein [A-1953, vol. 1, p. 136] has found polynomials of best approximation for $f_1(x) := (x-a)^{-1}$, $f_2(x) := (x^2 - a^2)^{-1}$, $a > 1$ and the errors $E_n(f_1) = \frac{1}{a^2-1} + \rho^n$,

$$E_{2n}(f_2) = \frac{1}{2a^2(a^2-1)} \rho^{2n}, \quad \rho := \frac{1}{a + \sqrt{a^2-1}}.$$

14.4. In Numerical Analysis, the convergence of an algorithm is called *linear* if the error e_m after the m-th computation satisfies $e_m \le C\alpha^m$ for some $C > 0$ and $0 < \alpha < 1$, and *quadratic,* if $e_m \le C\alpha^{2^m}$. Improving our Theorem 8.2, one can show the *quadratic convergence* of the Remez algorithm, if one makes additional assumptions about the functions, or modifies the algorithm. In Powell [A-1981, p. 106] the additional assumptions are that f, ϕ_i are continuously differentiable and that there are exactly $n+1$ alternation points for $f - P^*$. In particular, these assumptions imply that the endpoints a, b are alternation points. Under the same assumptions, Murnaghan and Wrench [1959] modify the Remez algorithm and obtain $T_+ := (t_i^+)$, $t_0^+ = a$, $t_n^+ = b$, from $T := (t_i)_0^n$, $t_0 = a$, $t_n = b$, by applying one step of the Newton algorithm for the non-linear system $f'(u_i) - P'(u_i) = 0$, $i = 1, \dots, n-1$ whose unknowns are the u_i.

14.5. Pinkus and Ziegler [1979] have proven some interesting properties of the zeros of error functions in L_p approximation, $1 < p \le \infty$. Assume that both $\{g_1, \dots, g_n\}$ and $\{g_1, \dots, g_n, \phi, \psi\}$ are Haar systems in $C[a,b]$. Let P_n and Q_n be the best L_p approximations from $\operatorname{span}\{g_0, \dots, g_n\}$ to ϕ and ψ respectively. Then the zeros of the error functions $\phi - P_n$ and $\psi - Q_n$ interlace on $[a,b]$. Slightly weaker results hold for the points of alternation of the error functions.

14.6. Pólya systems are not as special among Haar systems as they appear at first sight. According to a theorem of Karlin and Studden [A-1966, p. 379], a Haar system Φ has a representation as a Pólya system (11.3) if and only if all determinants (11.5) with $\lambda_k = k$ are > 0.

14.7. Pólya [1922] has introduced the functions (11.3) in the study of differential equations; they are a basis for the null space of the operator D of (11.4). Differential operators are called *disconjugate* if their null space is a Haar system. Zedek [1965] has shown that $(D + \lambda_1) \dots (D + \lambda_n)$ is disconjugate if $\lambda_k \in C^k$, $k = 1, \dots, n$.

Chapter 4. Properties of Polynomials

§ 1. Inequalities of Bernstein, Szegö and Markov

We begin with inequalities for the uniform norm. The simplest of these inequalities concerns trigonometric polynomials T_n of degree $\le n$, it is

$$\|T_n'\| \le n \|T_n\| . \tag{1.1}$$

This is best possible, for there is equality if $T_n = \sin nt$. Much older is Markov's inequality for algebraic polynomials P_n of degree $\le n$ on $[-1,1]$:

$$\|P_n'\| \le n^2 \|P_n\| . \tag{1.2}$$

We shall see many similar inequalities, in different norms; many of them have important applications. The proofs of the following two theorems are by v. Golitschek [1989].

Theorem 1.1 (Szegö [1928]). *For each trigonometric polynomial T_n of degree $\le n$, in the uniform norm on $\mathbb{T}$,*

$$T_n'(t)^2 + n^2 T_n(t)^2 \le n^2 \|T_n\|^2 , \quad t \in \mathbb{T} . \tag{1.3}$$

Proof. First, let $\|T_n\| < 1$. We can further assume that $t = 0$, $T_n'(0) \ge 0$. We define a real α, $|\alpha| < \pi/(2n)$ by $\sin n\alpha = T_n(0)$, and the trigonometric polynomial S_n by

$$S_n(x) := \sin n(x + \alpha) - T_n(x) .$$

At the points $t_k := -\alpha + \frac{(2k-1)\pi}{2n}$, $k = 0, \pm 1, \dots$, sign $S_n(t_k) = (-1)^{(k+1)}$, hence S_n has a unique zero c_k in each interval (t_k, t_{k+1}). Since $0 \in (t_0, t_1)$ and $S_n(0) = 0$, $c_0 = 0$. Also, $S_n(t_1) > 0$. The inequality $S_n'(0) \le 0$ would imply the existence of another zero of S_n in (t_0, t_1). Therefore $S_n'(0) > 0$, and

$$0 \le T_n'(0) = n \cos n\alpha - S_n'(0) < n \cos n\alpha = n\sqrt{1 - T_n(0)^2} .$$

This shows that for $t \in \mathbb{T}$,

$$T_n'(t)^2 + n^2 T_n(t)^2 \le n^2 .$$

For an arbitrary T_n, we take $\lambda > \|T_n\|$ and apply this relation to the polynomial T_n/λ. We obtain $T_n'(t)^2 + n^2 T_n(t)^2 \le \lambda^2 n^2$, and making $\lambda \to \|T_n\|$, we get (1.3). □

As a corollary, we have the inequality of Bernstein (1.1), and by iteration

$$\|T_n^{(k)}\| \le n^k \|T_n\|\,, \quad k = 1, 2, \dots . \tag{1.4}$$

Inequality (1.1) remains true for trigonometric polynomials T_n with complex coefficients. We select a real α in such a way that the $e^{i\alpha}T_n'$ attains the value $\|T_n'\|$, say, for $t = t_0$. Now $S_n(t) = \operatorname{Re}\{e^{i\alpha}T_n(t)\}$ is a polynomial with real coefficients for which $\|S_n\| \le \|T_n\|$ and $S_n'(t) = \operatorname{Re}\{e^{i\alpha}T_n'(t)\}$. Hence,

$$\|T_n'\| = e^{i\alpha}T_n'(t_0) = S_n'(t_0) \le n\|T_n\|\,. \tag{1.5}$$

Useful inequalities for algebraic polynomials follow.

Corollary 1.2. *For a polynomial P_n of degree $\le n$ we have on $[-1, 1]$,*

$$|P_n'(x)| \le \frac{n\|P_n\|}{\sqrt{1-x^2}}\,, \quad -1 < x < 1\,. \tag{1.6}$$

We can obtain this from (1.1) by the "standard" substitution $x = \cos t$. Since $T_n(t) := P_n(\cos t)$ is a trigonometric polynomial of degree $\le n$ and $\|T_n\| = \|P_n\|$, we have $|T_n'(t)| \le \|P_n\|n$. But $T_n'(t) = \pm P_n'(x)\sqrt{1-x^2}$.

Corollary 1.3. *For a polynomial $P(z) = \sum_0^n a_k z^k$ with complex coefficients for the uniform norm on the disk $|z| \le 1$,*

$$\|P_n'\| \le n\|P_n\|\,. \tag{1.7}$$

Proof. On the circumference $|z| = 1$, $z = e^{it}$, $t \in \mathbb{T}$. Then $P_n(z) = \sum_0^n a_k e^{ikt} = T_n(t)$ is a trigonometric polynomial with complex coefficients, and $P_n'(z) = -ie^{-it}T_n'(t)$. Thus, (1.7) follows from (1.5) and the maximum modulus principle. □

The inequality (1.6) deteriorates if x approaches the endpoints $-1, 1$ of the interval. One can give, however, a constant upper bound.

Theorem 1.4 (A.A. Markov). *For algebraic polynomials P_n of degree $\le n$ one has* (1.2).

It is clear that (1.2) cannot be improved. If P_n is the Chebyshev polynomial $C_n(x) = \cos nt$, $x = \cos t$, $0 \le t \le \pi$, then $\|C_n\| = 1$ and $C_n'(x) = n \sin nt/\sin t$ with $C_n'(1) = n^2$ for $x = 1$ (or $t = 0$). We also see that

$$|C_n'(x)| \le n^2\,, \quad -1 \le x \le 1\,.$$

This follows from the inequality $|\sin nt| \le n|\sin t|$, which can be easily proved by induction on n.

Proof. As in Theorem 1.1, it is enough to prove $\|P_n'\| \le n^2$, whenever $\|P_n\| < 1$. If $\xi := x_n = \cos\frac{\pi}{2n}$ (see (6.4) of Chapter 3) is the largest zero of C_n, then

$$\sqrt{1-\xi^2} = \sin\frac{\pi}{2n} \ge \frac{1}{n}\,. \tag{1.8}$$

(a) First let $|x| \le \xi$. Then (1.2) is a consequence of (1.6).

(b) In the case $\xi < |x| \le 1$, we can assume for example that $\xi < x \le 1$ and that $P'(x) > 0$. We have here $C_n(x) > 0$.

(b_1) Let $P_n(x) \ge C_n(x)$. Then for $T_n(t) := P_n(\cos t)$ and $x = \cos t$, Szegö's inequality implies

$$\begin{aligned} |P_n'(x)| = \left|\frac{T_n'(t)}{\sin t}\right| &\le \frac{n}{|\sin t|}\sqrt{1-T_n(t)^2} \\ &\le \frac{n}{|\sin t|}\sqrt{1-C_n(\cos t)^2} = n\left|\frac{\sin nt}{\sin t}\right| = |C_n'(x)| \le n^2\,. \end{aligned}$$

(b_2) Let now $P_n(x) < C_n(x)$. The polynomial $R := C_n - P_n$ has the sign of C_n at the extremal points $y_k = \cos\frac{k\pi}{n}$, $k = 0, 1, \dots, n$, of C_n. Consequently, $\operatorname{sign} R(y_k) = (-1)^k$. The interval $[-1, y_1)$ has n points y_k, hence at least $n-1$ zeros of R. Also, R changes sign from y_1 to x, hence it has a zero η, $y_1 < \eta < x$. It follows that all zeros of R' are in $[-1, \eta)$. On the interval $(\eta, +\infty)$, R' is of constant sign, and since $R(\eta) = 0$ and $R(x) > 0$, we have $R'(x) \ge 0$. This yields $0 \le P_n'(x) \le C_n'(x) \le n^2$. □

An iteration of (1.2) yields for $[-1, 1]$,

$$\|P_n^{(k)}\| \le n^2(n-1)^2 \dots (n-k+1)^2 \|P_n\|\,. \tag{1.9}$$

The factor on the right is not the best possible, but formula (1.9) is at least asymptotically correct for each fixed k and $n \to \infty$. Indeed, for the Chebyshev polynomials C_n one derives from the differential equation (6.5) of Chapter 3 by means of differentiation and evaluation at $x = 1$ the relation

$$C_n^{(k)}(1) = \frac{n^2-(k-1)^2}{2k-1} C_n^{(k-1)}(1)\,, \quad k = 1, 2, \dots .$$

Thus, by induction, $C_n^{(k)}(1) \ge \gamma_k n^{2k} = \gamma_k n^{2k}\|C_n\|$ for some $\gamma_k > 0$ and all large n. See also Note **12.1** (b).

§ 2. Polynomials on the Complex Plane and in Banach Spaces

If a bound for a polynomial P_n on $[-1, +1]$ is given, what can be said about $P_n(z)$ for z outside of this interval? Essential here is the transformation

$$z = \frac{1}{2}\left(w + \frac{1}{w}\right)\,, \quad |w| \le 1 \tag{2.1}$$

which maps the disk $|w| \le 1$ onto the complex plane and the circle $|w| = 1$ onto the interval $[-1,1]$. Its inverse is $w = z \pm (z^2-1)^{1/2}$. We see that there is a unique w which satisfies $|w| < 1$ if $z \notin [-1,+1]$. Otherwise, we have $z = \cos t$ with real t, and there are two values $w = e^{it}$, e^{-it}.

In the real form, (2.1) can be written

$$x = \frac{1}{2}(\rho + \rho^{-1})\cos\phi\,, \quad y = \frac{1}{2}(\rho - \rho^{-1})\sin\phi\,, \tag{2.2}$$

where $z = x + iy$, $w = \rho e^{i\phi}$. The equations (2.2) describe a curve E_ρ when ρ is fixed. For $\rho \ne 1$, this is the ellipse $x^2/a^2 + y^2/b^2 = 1$ with the semi-axes $a = \frac{1}{2}(\rho+\rho^{-1})$ and $b = \frac{1}{2}|\rho-\rho^{-1}|$, and the foci ± 1. The ellipse $E_{\rho^{-1}}$ coincides with E_ρ. For $\rho = 1$, E_ρ is the interval $[-1,+1]$. There is exactly one ellipse E_ρ, $\rho > 1$ passing through z; ρ is found from the equation $\rho+\rho^{-1} = |z-1|+|z+1|$.

Assume now that we have

$$|P_n(x)| \le \Omega(x)\,, \quad x \in [-1,+1]\,, \tag{2.3}$$

where $\Omega(z) = M\prod_{k=1}^N |z-c_k|^{-s_k}$. Here s_k are real constants, and c_k are complex constants. We would like to derive from (2.3) an estimate for $P_n(z)$ for arbitrary complex z.

We define w_k for $k = 1,\dots,N$ by $c_k = \frac{1}{2}(w_k + w_k^{-1})$, $|w_k| \le 1$; if $c_k \in [-1,+1]$, we fix one of the two possible choices of w_k. From (2.1) we obtain

$$|z - c_k| = |(2w_k)^{-1}(1-w_k w)(1-w_k w^{-1})|\,. \tag{2.4}$$

If $|w| = 1$, $w^{-1} = \bar{w}$; hence

$$|z - c_k| = |(2w_k)^{-1}(1-w_k w)(1-\bar{w}_k w)|\,. \tag{2.5}$$

Theorem 2.1 (Kemperman and Lorentz [1979]). *If (2.3) holds for a polynomial P_n of degree $\le n$, then for all complex z,*

$$|P_n(z)| \le M|w|^{-n}\prod_{k=1}^N |(2w_k)^{-1}(1-w_k w)(1-\bar{w}_k w)|^{-s_k}\,. \tag{2.6}$$

An equivalent inequality is

$$|P_n(z)| \le \Omega(z)|w|^{-n}\prod_{k=1}^N \left|\frac{w^{-1}(w-w_k)}{1-\bar{w}_k w}\right|^{s_k}\,. \tag{2.7}$$

Proof. From (2.4) it follows that (2.6) and (2.7) are equivalent. To prove the first inequality, we remark that

$$F(w) := \prod_{k=1}^N \{(2|w_k|)^{-1}(1-w_k w)(1-\bar{w}_k w)\}^{s_k}$$

is analytic for $|w| \le 1$, except for possible singularities at w_k, $\bar{w}_k$ when $|w_k| = 1$. With the polynomial

$$Q(w) := w^n P_n\left(\frac{w+w^{-1}}{2}\right),$$

inequality (2.6) becomes $|Q(w)F(w)| \le M$ for $|w| \le 1$. By the maximum modulus theorem, this inequality holds since it holds for $|w| = 1$ because of (2.3), and since $Q(w)F(w)$ is analytic in $|w| < 1$ and continuous on $|w| = 1$. □

The simplest special case of Theorem 2.1 is when $\Omega(z) \equiv M$:

Theorem 2.2 (Bernstein). *If a polynomial P_n with complex coefficients satisfies $|P_n(x)| \le M$ for $-1 \le x \le 1$, then*

$$|P_n(z)| \le M\rho^n\,, \quad z \in E_\rho\,, \quad \rho > 1\,. \tag{2.8}$$

Another useful inequality compares arbitrary real polynomials $P_n \in \mathcal{P}_n$ with the Chebyshev polynomials. For real x, it is somewhat stronger than (2.8).

Proposition 2.3. *If a polynomial P_n of degree $\le n$, satisfies $|P_n(x)| \le M$, $x \in [-1,1]$, then*

$$|P_n(x)| \le M|C_n(x)|\,, \quad x \notin [-1,1]\,. \tag{2.9}$$

Proof. For any $M_0 > M$, the polynomials $Q := M_0 C_n \pm P_n$ has n sign changes and hence n zeros in $(-1,1)$. Therefore, Q has no additional sign changes. Since $|M_0 C_n(\pm 1)| > |P_n(\pm 1)|$, Q has the sign of C_n outside $(-1,1)$. This gives (2.9) with M replaced by M_0. Letting $M_0 \to M$, we obtain (2.9). □

Here are simple corollaries. If $|P_n(x)| \le M$ on $I := [-1,1]$, then on an interval $J := [-\lambda,\lambda]$, $\lambda > 1$, concentric to I we have $|C_n(x)| \le (\sqrt{\lambda^2-1}+\lambda)^n \le 2^n\lambda^n$, hence $|P_n(x)| \le M|J|^n$, $x \in J$. For arbitrary intervals I, J, we get from this

$$\|P_n\|_\infty(J) \le C_n(|J|/|I|)^n \|P_n\|_\infty(I)\,, \quad \text{if } I \subset J\,. \tag{2.10}$$

We shall next discuss Bernstein type inequalities for Banach norms. They follow from the following result about the order relation $f \prec g$. According to (2.5) of Chapter 2, the relation $f \prec g$ for two integrable functions on $\mathbb{T}$ holds if and only if, for each measurable set $A \subset \mathbb{T}$, there is another set $B \subset \mathbb{T}$ of equal measure so that $\int_A |f|\,dt \le \int_B |g|\,dt$.

Theorem 2.4 (Lorentz [1984]). *For each trigonometric polynomial T_n of degree $\le n$ with complex coefficients,*

$$T_n' \prec nT_n\,. \tag{2.11}$$

Proof. We write sign $\beta = \beta/|\beta|$ if $\beta \neq 0$, sign $0 = 0$. For a given set $A \subset \mathbb{T}$, with the characteristic function χ_A, we define

$$Q(t) = \int_{\mathbb{T}} \chi_A(s) T_n(s+t) \text{ sign } T_n'(s)\, ds\,.$$

This is a trigonometric polynomial of degree $\le n$. Let $t_0 \in \mathbb{T}$ be such that $|Q(t_0)| = \|Q\|_\infty$. Using Bernstein's inequality (1.1) for the value $Q(0)$, we obtain

$$\begin{aligned}\int_A |T_n'(s)|\, ds &= \int_{\mathbb{T}} \chi_A(s)|T_n'(s)|\, ds = Q'(0)\\ &\le n|Q(t_0)| \le n \int_{\mathbb{T}} \chi_A(s)|T_n(s+t_0)|\, ds\\ &= n\int_B |T_n(s)|\, ds\,,\end{aligned}$$

where B is the $-t_0$-translation of A. Since $|B| = |A|$, we have (2.12). □

An iteration of (2.10) shows that $T_n^{(r)} \prec n^r T_n$ for $r = 1, 2, \ldots$. In a rearrangement-invariant space X on $\mathbb{T}$, $f \prec g$ implies $\|f\|_X \le \|g\|_X$. It follows:

Theorem 2.5. *For each rearrangement invariant space X on $\mathbb{T}$, and for $T_n \in \mathcal{T}_n$,*

(2.12) $$\|T_n^{(r)}\|_X \le n^r \|T_n\|_X\,, \quad r = 1, 2, \ldots\,.$$

In particular,

(2.13) $$\|T_n^{(r)}\|_p \le n^r \|T_n\|_p\,, \quad 1 \le p \le \infty$$

(first proved by Zygmund). The case $0 < p < 1$, which we discuss in § 3, is more difficult.

Next we discuss the useful inequalities of Nikolskii [1951], which compare the norms of P_n or of T_n in different metrics. In the inequalities of the type $C_1\|P_n\|_q \le \|P_n\|_p \le C_2\|P_n\|_q$, $q < p \le \infty$, we want to find out how the constants depend on n. Of interest is only the second relation – the first is a consequence of Hölder's inequality. For P_n on $[a, b]$, the inequality is best formulated in terms of the norms

$$\|f\|_p^*[a,b] := \|f\|_p^* := \left(\frac{1}{b-a}\int_a^b |f|^p\, dx\right)^{1/p}.$$

Theorem 2.6. *For $0 < q \le p \le \infty$ and polynomials $P_n \in \mathcal{P}_n$, $n = 1, 2, \ldots$, on $[a, b]$, and for trigonometric polynomials $T_n \in \mathcal{T}_n$ on $\mathbb{T}$,*

(2.14) $$\|P_n\|_p^* \le \left(2(q+1)n^2\right)^{1/q-1/p}\|P_n\|_q^*\,,$$

(2.15) $$\|T_n\|_p \le \left(\frac{2nr+1}{2\pi}\right)^{1/q-1/p}\|T_n\|_q\,,$$

where $r := r(q)$ is the least integer $\ge q/2$.

Proof. (a) We can assume that the maximum $\|P_n\|_\infty$ of $|P_n(x)|$ is attained at some point $x_0 \le \frac{1}{2}(a+b)$. Then the interval $[x_0, b]$ contains $I = [x_0, x_0 + \frac{b-a}{2n^2}]$. From $|P_n(x_0) - P_n(x)| \le (x - x_0)\|P_n'\|_\infty$, $x \ge x_0$ and the inequality $\|P_n'\|_\infty \le \frac{2n^2}{b-a}\|P_n\|_\infty$ (which follows from (1.2) by a linear substitution) we have $|P_n(x)| \ge (1 - \frac{2n^2}{b-a}(x - x_0))\|P_n\|_\infty \ge 0$ for $x \in I$. Raising this to the q-th power and integrating,

$$\int_{x_0}^b |P_n|^q\, dx \ge \|P_n\|_\infty^q \left(\frac{2n^2}{b-a}(q+1)\right)^{-1}.$$

This yields $\|P_n\|_\infty \le (\frac{2n^2}{b-a}(q+1))^{1/q}\|P_n\|_q$, and since $\|P_n\|_p \le \|P_n\|_\infty^{1-q/p}\|P_n\|_q^{q/p}$, we obtain (2.14).

(b) To establish (2.15), we consider the trigonometric polynomial T_n^r of degree $m := rn$ and represent it by the identity

$$T_n(t)^r = \frac{1}{\pi}\int_0^{2\pi} T_n(x)^r D_m(x-t)\, dx\,,$$

where $D_m(t) := \frac{1}{2} + \sum_{j=1}^m \cos jt$ is the Dirichlet kernel. This implies

$$\pi\|T_n\|_\infty^r = \pi\|T_n^r\|_\infty \le \|T_n\|_\infty^{r-q/2}\int_0^{2\pi} |T_n(x)|^{q/2}|D_m(x-t)|\, dx\,.$$

Using the identity $\|D_m\|_2 = \sqrt{(m+1/2)\pi}$ and the Cauchy-Schwarz inequality we obtain

$$\pi^2\|T_n\|_\infty^q \le (m+1/2)\pi\|T_n\|_q^q$$

and thus

$$\|T_n\|_\infty \le \left(\frac{2nr+1}{2\pi}\right)^{1/q}\|T_n\|_q\,,$$

which establishes (2.15) for $p = \infty$. We get (2.15) for all $p > q$ since we can use the inequality $\|T_n\|_p \le \|T_n\|_\infty^{1-q/p}\|T_n\|_q^{q/p}$. □

It is even simpler to compare the norms of polynomials P_n and their derivatives on $I = [a, b]$ when the degrees n are bounded, but the length $\delta := b - a$ of I varies.

Theorem 2.7. *Let r and p_0 be given. Then for all polynomials $P \in \mathcal{P}_r$ on I, all $p, q \ge p_0$, $k = 1, \ldots, r$, there is a constant $C = C(p_0, r)$ such that*

(2.16) $$\|P\|_q(I) \le C|I|^{1/q-1/p}\|P\|_p(I)\,,$$

(2.17) $$\|P^{(k)}\|_q(I) \le C|I|^{-k+1/q-1/p}\|P\|_p(I)\,.$$

Proof. We can assume that $I = [0, \delta]$. The spaces $\mathcal{P}_r$ on $A := [0, 1]$ equipped with the (quasi-) norm $\|\cdot\|_p$ are complete metric spaces and have therefore equivalent norms (contained between $\|\cdot\|_{p_0}$ and $\|\cdot\|_\infty$). Thus, $\|P\|_q(A) \sim \|P\|_p(A)$, $P \in \mathcal{P}_r$, with constants not exceeding some $C(p_0, r)$ in equivalencies. On the other hand, if $Q \in \mathcal{P}_r$ is defined by $Q(u) := P(\delta u)$, the substitution $x = \delta u$ reveals that $\|P\|_q(I) = \delta^{1/q}\|Q\|_q(A)$, hence

$$\|P\|_q(I) \le C\delta^{1/q}\|Q\|_p(A) = C\delta^{1/q-1/p}\|P\|_p(I)\,.$$

In the same way, by Markov's inequality, and since $P^{(k)}(\delta u) = \delta^{-k}Q^{(k)}(u)$,

$$\begin{aligned}\|P^{(k)}\|_q(I) &= \delta^{-k+1/q}\|Q^{(k)}\|_q(A) \le C\delta^{-k+1/q}\|Q^{(k)}\|_\infty(A)\\ &\le C\delta^{-k+1/q}\|Q\|_\infty(A) \le C\delta^{-k+1/q}\|Q\|_p(A)\\ &= C\delta^{-k+1/q-1/p}\|P\|_p(I)\,.\end{aligned}$$

□

Obviously, this proof works for all linear spaces for which the norm of the dilation $Q(x) \to Q(\delta x)$ is known.

§ 3. Bernstein Inequalities in L_p, $0 < p < 1$

Methods of complex variables are needed to obtain these inequalities. For $p \to +0$, the semi-norm of a function $f(t)$, $t \in \mathbb{T}$,

$$\|f\|_p = \left\{\frac{1}{2\pi}\int_{\mathbb{T}} |f(t)|^p\, dt\right\}^{1/p}$$

converges (see Hardy, Littlewood and Pólya [B-1968]) to

$$\|f\|_0 = \exp\left(\frac{1}{2\pi}\int_{\mathbb{T}} \log|f(t)|\, dt\right)\,; \tag{3.1}$$

this defines the space L_0.

We shall prove that

$$\left\|\frac{1}{n}T_n'\right\|_p \le \|T_n\|_p\,, \quad 0 \le p \le 1 \tag{3.2}$$

for each trigonometric polynomial T_n of degree $\le n$ with real or complex coefficients. We shall actually prove this for all p, $0 \le p \le +\infty$, obtaining in this way a new approach to inequalities (1.3), (2.13) and new insights.

Several authors proved (3.2) with a constant factor on the right independent of n; Arestov [1981] was the first to establish (3.2) as it stands. We shall follow the proof of v. Golitschek and Lorentz [1989].

We shall use elementary facts about subharmonic functions (see Hille [B-1962]) and Jensen's formula for a function $f(z)$ analytic in $|z| \le 1$ with $f(0) \ne 0$:

$$\frac{1}{2\pi}\int_{\mathbb{T}} \log|f(e^{it})|\, dt = \log|f(0)| + \sum{}' \log\frac{1}{|z_k|}\,, \tag{3.3}$$

the sum $\sum'$ extended over all zeros z_k of f in $|z| < 1$. As an example, we apply (3.3) to the function $f(z) = z + e^{i\alpha}w$ with a complex w and a real α and obtain

$$\log^+|w| = \frac{1}{2\pi}\int_{\mathbb{T}} \log|w + e^{is}|\, ds = \frac{1}{2\pi}\int_{\mathbb{T}} \log|w + e^{i\alpha}e^{is}|\, ds\,. \tag{3.4}$$

We shall also use a form of Rolle's theorem:

Lemma 3.1. *If a real analytic function f has $2p$ zeros on $\mathbb{T}$ (counting multiplicities), then for any real A, B, the function $g = Af + Bf'$ has at least $2p$ zeros on $\mathbb{T}$.*

Proof. We can assume that $B \ne 0$. If t is a zero of f of multiplicity l, then it is a zero of g of multiplicity $l-1$. Let now a, b be two zeros of f, with no zeros of f on the arc a, b subtended in the clockwise direction. The function f does not change sign on this arc, let for example $f(t)$ be > 0 there. If $f(a) = \cdots = f^{(r)}(a) = 0$ but $f^{(r+1)}(a) \ne 0$, then $f(a+h) = \alpha h^{r+1} + O(h^{r+2})$ with $\alpha > 0$, and $f'(a+h) = (r+1)\alpha h^r + O(h^{r+1})$. Therefore, $f(a+h) = o(f'(a+h))$ and $f'(a+h) > 0$ for small $h > 0$. It follows that $g(a+h)$ has the sign of B. In the same way, $g(b-h)$ has the sign of $-B$. Thus, g changes sign on the arc a, b. □

To obtain inequalities more general than (3.2), let A, B be two real numbers satisfying $A^2 + B^2 = 1$, so that

$$|A - iB| = 1\,. \tag{3.5}$$

We define a linear operator, which maps the space $\mathcal{T}_n$ of all trigonometric polynomials T_n with complex coefficients onto itself by

$$T_n \to S_n = AT_n + B\frac{T_n'}{n}\,. \tag{3.6}$$

Theorem 3.2. *For all $T_n \in \mathcal{T}_n$,*

$$\frac{1}{2\pi}\int_{\mathbb{T}} \log|S_n(t)|\, dt \le \frac{1}{2\pi}\int_{\mathbb{T}} \log|T_n(t)|\, dt\,. \tag{3.7}$$

Proof. It is enough to prove (3.7) for $T_n(t) = \sum_{-n}^{n} c_k e^{ikt}$ with $c_n \ne 0$. In fact, every trigonometric polynomial of degree $\le n$ is a limit of such polynomials. We can then normalize T_n by $c_n = 1$. With each such T_n, we associate the algebraic polynomial of degree $2n$

$$P(z) := P(T_n; z) = \sum_{k=0}^{2n} c_{-n+k} z^k\,. \tag{3.8}$$

An alternate definition is $P(e^{it}) = e^{int}T_n(t)$. Given P, there is a unique T_n for which (3.8) holds. Differentiating, we obtain

$$T_n'(t) = ie^{-int}\left[-nP(z) + zP'(z)\right], \quad z = e^{it}.$$

From this relation, it is easy to derive that $t \in \mathbb{T}$ is a zero of T_n of order l exactly when $z = e^{it}$ is a zero of P of order l. Putting $R(z) = P(S_n; z)$ we have $R(e^{it}) = e^{int}S_n(t)$ and

$$R(z) = (A - iB)P(z) + \frac{iBz}{n}P'(z). \tag{3.9}$$

We see that (3.7) is equivalent to

$$\frac{1}{2\pi}\int_{\mathbb{T}} \log|R(e^{it})|\,dt \le \frac{1}{2\pi}\int_{\mathbb{T}} \log|P(e^{it})|\,dt.$$

The difference of these two integrals is equal to

$$\begin{aligned} F(Z) := F_n(z_1, \dots, z_{2n}) :&= \frac{1}{2\pi}\int_{\mathbb{T}} \log\left|A - iB + \frac{iBe^{it}}{n}\frac{P'(e^{it})}{P(e^{it})}\right| dt \\ &= \frac{1}{2\pi}\int_{\mathbb{T}} \log\left|A - iB + \frac{iBe^{it}}{n}\sum_{k=1}^{2n}\frac{1}{e^{it} - z_k}\right| dt \end{aligned} \tag{3.10}$$

where $Z := (z_1, \dots, z_{2n}) \in \mathbb{C}^{2n}$ and the z_k are the zeros of P. The function F is a continuous function on $\mathbb{C}^{2n}$ and we want to show that $F(Z) \le 0$, for all $Z \in \mathbb{C}^{2n}$.

Now if f is a univariate analytic function in some region (that may contain the point ∞), then $\log|f|$ is subharmonic in this region; it takes the value $-\infty$ at the zeros of f. Sums with positive coefficients and integrals with respect to a parameter (for a positive measure) are also subharmonic. Thus, if z_j, $j \ne k$, are fixed, $\phi_k(z) := F(z_1, \dots, z_{k-1}, z, z_{k+1}, \dots, z_{2n})$ is a subharmonic function in each of the regions $|z| < 1$ and $|z| > 1$.

We claim that F assumes its maximum at a point $Z^* = (z_1^*, \dots, z_{2n}^*)$ with $|z_k^*| = 1$, $k = 1, \dots, 2n$. Indeed, for each $Z \in \mathbb{C}^{2n}$, there is a point $\widetilde{Z}$ with $|\tilde{z}_k| = 1$, $k = 1, \dots, 2n$, with $F(Z) \le F(\widetilde{Z})$. To construct $\widetilde{Z}$, we suppose that for some k, we have $|z_k| \ne 1$, say for example $|z_k| > 1$. Then, the function ϕ_k assumes its maximum on the boundary of the region $|z| \ge 1$ at some point z_k' with $|z_k'| = 1$. Hence $|F(Z)| \le |F(Z')|$ for the point Z' with $z_j' := z_j$, $j \ne k$. Repeating this for Z' and so on, we arrive at $\widetilde{Z}$. It remains to use the continuity of $F(Z)$ for $|z_k| = 1$, $k = 1, \dots, 2n$.

Now, we write $z_k^* = e^{it_k}$ and let T_n^* be the trigonometric polynomial of degree n with zeros at t_k, $k = 1, \dots, 2n$, and coefficient $c_n = 1$. Then T_n^* is a real polynomial and by Lemma 3.1 all zeros of S_n^* are also real. But this is in turn equivalent to the assumption that all zeros of R^* lie on $|z| = 1$.

From (3.9) and (3.5) we see that $|P(0)| = |R(0)| \ne 0$. A direct application of Jensen's inequality yields now

$$F(Z^*) = \frac{1}{2\pi}\int_{\mathbb{T}} \log|R^*|\,dt - \frac{1}{2\pi}\int_{\mathbb{T}} \log|P^*|\,dt = 0$$

as desired. □

This argument shows also that one has equality in (3.7) if all zeros of T_n are real.

Theorem 3.3. *For all $T_n \in \mathcal{T}_n$,*

$$\frac{1}{2\pi}\int_{\mathbb{T}} \log^+|S_n(t)|\,dt \le \frac{1}{2\pi}\int_{\mathbb{T}} \log^+|T_n(t)|\,dt. \tag{3.11}$$

Proof. To replace log by $\log^+$, we apply (3.7) to the polynomials

$$T_n^*(t) := T_n^*(t,s) := T_n(t) + e^{is}e^{int},$$

which depend on the real parameter s. The linear map (3.6) gives the corresponding

$$S_n^*(t) := S_n^*(t,s) = S_n(t) + e^{is}e^{int}e^{i\alpha}, \quad A + iB = e^{i\alpha}.$$

By (3.7),

$$\int_{\mathbb{T}} \log|S_n(t) + e^{is}e^{i\alpha}e^{int}|\,dt \le \int_{\mathbb{T}} \log|T_n(t) + e^{is}e^{int}|\,dt.$$

Then, by means of (3.4),

$$\begin{aligned} \int_{\mathbb{T}} \log^+|S_n(t)|\,dt &= \int_{\mathbb{T}} dt \frac{1}{2\pi}\int_{\mathbb{T}} \log|S_n(t) + e^{is}e^{i\alpha}e^{int}|\,ds \\ &\le \int_{\mathbb{T}} dt \frac{1}{2\pi}\int_{\mathbb{T}} \log|T_n(t) + e^{is}e^{int}|\,ds \\ &= \int_{\mathbb{T}} \log^+|T_n(t)|\,dt. \end{aligned}$$ □

In order to get from $\log^+$ to the function $(\cdot)^p$ we can use the formula

$$u^p = p^2\int_0^\infty s^{p-1}\log^+\frac{u}{s}\,ds, \quad p > 0.$$

We can even slightly generalize it. Let $\Phi(u)$ and $\Psi(u) = u\Phi'(u)$ be continuous positive increasing functions defined for $u \ge 0$ with $\Phi(0) = \Psi(0) = 0$. The functions u^p, $\log^+ u$, $\log(1 + u^p)$, $p > 0$, are examples. Then

$$\Phi(u) = \int_0^{+\infty} \log^+ \frac{u}{s}\, d\Psi(s)\,. \tag{3.12}$$

Indeed, for $0 < v < u$,

$$\begin{aligned}\Phi(u) - \Phi(v) &= \int_v^u \Phi'(s)\,ds = -\int_v^u s\Phi'(s)\,d\log\frac{u}{s}\\ &= v\Phi'(v)\log\frac{u}{v} + \int_v^u \log\frac{u}{s}\,d\Psi(s)\,.\end{aligned} \tag{3.13}$$

The last integral is majorized by $\Phi(u) - \Phi(v)$, which has the limit $\Phi(u)$ for $v \to 0$. Hence the integral $\int_0^u \log(u/s)\,d\Psi(s)$ converges. Also the first term on the right has a limit $C \geq 0$. The assumption $C > 0$ leads to a contradiction, since then $\Phi'(v) \geq \text{Const}\,/v\log(u/v)$ and $\int_0^u \Phi'(v)\,dv$ diverges. Letting $v \to 0$ in (3.13) we obtain (3.12). The following theorem is due to Arestov for $A = 0$, $B = 1$, to v. Golitschek and Lorentz in general.

Theorem 3.4. *For each function Φ of the described type and for $T_n \in \mathcal{T}_n$,*

$$\int_{\mathbb{T}} \Phi(|S_n(t)|)\,dt \leq \int_{\mathbb{T}} \Phi(T_n(t))\,dt\,. \tag{3.14}$$

Proof. By means of (3.11) and (3.12),

$$\begin{aligned}\int_{\mathbb{T}} \Phi(|S_n(t)|)\,dt &= \int_0^\infty \int_{\mathbb{T}} \log^+\left|\frac{S_n(t)}{s}\right| dt\,d\Psi(s)\\ &\leq \int_0^\infty \int_{\mathbb{T}} \log^+\left|\frac{T_n(t)}{s}\right| dt\,d\Psi(s)\\ &= \int_{\mathbb{T}} \Phi(|T_n(t)|)\,dt\,.\end{aligned} \tag{3.15}$$

□

The special case $\Phi(u) = u^p$ of (3.14) reads

$$\left\|\cos\alpha T_n + \sin\alpha\frac{T_n'}{n}\right\|_p \leq \|T_n\|_p\,, \tag{3.16}$$

for all real α. This is also true for the limiting case $p = \infty$, when the inequality takes the form, for real or complex polynomials T_n

$$\left|T_n(t)\cos\alpha + \frac{1}{n}T_n'(t)\sin\alpha\right| \leq \|T_n\|_\infty\,, \quad t \in \mathbb{T}\,. \tag{3.17}$$

The maximum of $a\cos\alpha + b\sin\alpha$ for real constants a, b is $\sqrt{a^2+b^2}$. From this point of view, the inequality (1.3) of Szegö is nothing but the real variable form of inequality (3.17).

Theorem 3.5. *Let $B \neq 0$, and let Ψ be strictly increasing. Equality holds in (3.14) if and only if $T_n(t)$ is of the form $C_1e^{-int} + C_2e^{int}$.*

Proof. We have equality in (3.15) if and only if T_n satisfies

$$\int_{\mathbb{T}} \log^+\left|\frac{S_n(t)}{s}\right| dt = \int_{\mathbb{T}} \log^+\left|\frac{T_n(t)}{s}\right| dt$$

for all $s > 0$. This set of the T_n does not depend on the function Φ. We can assume that $\Phi(u) = u^2$ in (3.14). If $T_n(t) = \sum_{-n}^n c_ke^{ikt}$, then we have $\sum_{-n}^n |A + iBk/n|^2|c_k|^2 = \sum_{-n}^n |c_k|^2$, which immediately gives the desired result. □

§ 4. Polynomials with Positive Coefficients in x, $1-x$

For some special classes of polynomials, the estimates of Bernstein (1.6) and of Markov (1.2) can be improved. Inequalities (4.6) and (4.7) (Lorentz [1963]) below are of this type. They are valid for the derivatives $P^{(r)}$ of all orders, but lacks exact constants. Later Scheick [1966, 1972] proved exact estimates for $r = 1, 2$ for the inequalities of Markov type.

Since the polynomials $q_{n,k}(x) := x^k(1-x)^{n-k}$, $0 \leq k \leq n$ are linearly independent, each polynomial P of degree $\leq n$ can be uniquely written in the form $P = \sum_{k=0}^n a_kq_{n,k}$. If all $a_k \geq 0$, then P is a *polynomial with positive coefficients in x and $1-x$*. We begin with some examples and simple facts.

1. The minimal n for which a polynomial P is a polynomial with positive coefficients of degree n may be larger than the ordinary degree of P.

2. Sums and products of polynomials with positive coefficients are again polynomials of this type.

3. The Bernstein polynomials of a function $f(x) \geq 0$, $0 \leq x \leq 1$ are polynomials with positive coefficients.

4. If P is a real polynomial and all its zeros lie in the region $|z| \geq 1$, then either P or $-P$ is a polynomial with positive coefficients in $1+x$ and $1-x$. Indeed, $P(x)$ is a constant multiple of a product of terms of types $x+\alpha$, or $\alpha - x$ with $\alpha \geq 1$ and $(x+\alpha)^2 + \beta^2$ with $\alpha^2 + \beta^2 \geq 1$; all of which are polynomials with positive coefficients in $1+x$, $1-x$.

5. The derivative of a polynomial with positive coefficients is not necessarily a polynomial of this type. On the contrary, each real polynomial is the derivative of a polynomial with positive coefficients. This follows from the fact that for each real polynomial P_n, $P_n(x)+M = P_n(x)+M[x+(1-x)]^n$ is a polynomial with positive coefficients if the constant M is large enough.

As a consequence, inequalities of the Bernstein or Markov type for $P_n^{(r)}$, $r > 1$ do not automatically follow from those for the first derivative P_n', but have to be proved separately.

In what follows, we often extend the definition of $q_{n,k}$, $0 \leq k \leq n$ to non-integer k, n. We put $q_{n,k}(x) = 0$ for $x \notin [0,1]$. On the way to Theorem 4.2, we prove with Scheick [1972]:

Lemma 4.1. *Let* $\alpha := \alpha(x) = 1-2x$, $t := t(x) := x + \frac{\alpha}{n}$, *then for* $n = 2, 3, \ldots$ *and* $0 \le x \le \frac{1}{2}$

$$q'_{n,k}(x) \le enq_{n,k}(t)\ ; \quad q'_{n,k}(x) \ge -2nq_{n,k}(x)\,. \tag{4.1}$$

Proof. Clearly $x \le t \le \frac{1}{2}$, $t = (1 - \frac{2}{n})x + \frac{1}{n}$. The second inequality (4.1) is easy: from $q'_{n,k}(x) = (k - nx)x^{k-1}(1-x)^{n-k-1}$, we derive

$$\frac{1}{n}\frac{q'_{n,k}(x)}{q_{n,k}(x)} = \frac{(k/n) - x}{x(1-x)} \ge 4\left(\frac{k}{n} - \frac{1}{2}\right) \ge -2\,.$$

To prove the first inequality we begin with the formula

$$r_{n,k}(x) := \frac{q'_{n,k}(x)}{nq_{n,k}(t)} = \frac{(k/n) - x}{t(1-t)}\left(\frac{x}{t}\right)^{k-1}\left(\frac{1-x}{1-t}\right)^{n-k-1}. \tag{4.2}$$

The inequality $1 + u < e^u$, $u \ne 0$ yields

$$\frac{x}{t} = 1 - \frac{\alpha}{nt} \le \exp\left(-\frac{\alpha}{nt}\right),$$
$$\frac{1-x}{1-t} = 1 + \frac{\alpha}{n(1-t)} \le \exp\left(\frac{\alpha}{n(1-t)}\right)\ ;$$

for $r_{n,k}(x)$ we obtain the upper bound

$$U_k(x) := w\exp\left[\alpha\left(-\frac{k-1}{nt} + \frac{n-k-1}{n(1-t)}\right)\right] = we^{-\alpha w}e^{T+S},$$

where

$$w = \frac{(k/n) - x}{t(1-t)}, \quad T = \frac{\alpha(\alpha+1)}{nt}, \quad S = \frac{\alpha(\alpha-1)}{n(1-t)}\ ;$$

S and T do not depend on k.

We want to prove that $U_k(x) \le e$ if $0 \le x \le \frac{1}{2}$, $0 \le k \le n$. We fix x and study $U_k(x)$ as a function of a continuous variable k. Since

$$\frac{d}{dk}(we^{-\alpha w}) = e^{-\alpha w}(1 - \alpha w)\frac{1}{nt(1-t)},$$

the function $U_k(x)$ has a maximum at $k = k^*(x)$, where $k^*(x)$ is defined by $w = \frac{1}{\alpha}$, that is, by $\frac{k^*(x)}{n} = x + \frac{t(1-t)}{\alpha}$. We consider two cases.

Case I: $\frac{1}{3} \le x \le \frac{1}{2}$. Since

$$\frac{d}{dx}\left(\frac{k^*}{n}\right) = 1 + \left(1 - \frac{2}{n}\right)\frac{1-2t}{\alpha} + \frac{2}{\alpha^2}t(1-t) > 0,$$

$k^*(x)$ increases with x on $[0, \frac{1}{2}]$. For $x = \frac{1}{3}$,

$$k^*\left(\frac{1}{3}\right) = \frac{n}{3} + \frac{2}{3}\frac{(n+1)(n-\frac{1}{2})}{n} \ge n\,.$$

It follows that $k^*(x) \ge n$ on $[\frac{1}{3}, \frac{1}{2}]$ and so $U_{k^*} \le U_n$ on this interval. To estimate the right-hand side of

$$U_n(x) = \frac{1-x}{t(1-t)}\exp\left[\alpha\left(-\frac{n-1}{nt} - \frac{1}{n(1-t)}\right)\right] \le \frac{1}{t}\exp\left(-\frac{n-1}{n}\frac{\alpha}{t}\right)$$

we note that its derivative with respect to x

$$= \frac{1}{t^2}\exp\left(-\frac{n-1}{n}\frac{\alpha}{t}\right)\left[1 + \frac{n-1}{n}\left(1 - \frac{2}{n}\right)\frac{\alpha}{t}\right] > 0,$$

so that $U_n(x) \le U_n(\frac{1}{2}) = 2$, and we have established $U_k(x) \le 2$ in Case I.

Case II: $0 \le x \le \frac{1}{3}$. We can express T and S in terms of t:

$$T = At + B + \frac{2(n-1)}{(n-2)^2t}, \quad S = Ct + D + \frac{2(n-1)}{(n-2)^2(1-t)},$$

where A, B, C, D are constants (depending on n). Now

$$\log U_{k^*} = \log\frac{1}{\alpha} - 1 + T + S$$

and each of the functions on the right has a non-negative second derivative with respect to x (or to t). Hence, $\log U_{k^*}(x)$ as well as $U_{k^*}(x)$ attain their maximal values on $[0, \frac{1}{3}]$ at one of the endpoints of this interval. One computes easily that $U_{k^*}(0) = e$ and that for $n = 1, 2$

$$U_{k^*(1/3)} = 3\exp\left(-1 + \frac{4}{3(n+1)} - \frac{1}{3}\frac{1}{n - (1/2)}\right) < e\,.$$

The inequality is valid also for $n > 2$, since then the expression in the bracket is decreasing. □

By means of this lemma, we prove

Theorem 4.2 (Scheick). *For a polynomial P_n with positive coefficients on* $[0, 1]$, *and the uniform norm on this interval,*

$$\|P'_n\| \le en\|P\|\ ; \tag{4.3}$$

the constant e is asymptotically the best possible.

Proof. If $P_n = \sum_0^n a_k q_{n,k}$ with $a_k \ge 0$, then by (4.1),

$$-2n\|P_n\| \le -2nP_n(x) \le P'_n(x) \le enP_n(t) \le en\|P_n\|$$

for $0 \le x \le \frac{1}{2}$. Replacing x by $1 - x$, we obtain the complete result.

The last statement of the theorem follows from

$$\lim_{n\to\infty}\frac{\|q'_{n,1}\|}{n\|q_{n,1}\|} = e\,. \qquad \square$$

An interesting special case (we use **4** and a linear transformation of $[0,1]$ onto $[-1,1]$) is given by

Theorem 4.3. *Let $P_n(z)$ be a polynomial with real coefficients and with zeros outside of the circle $|z|<1$. Then*

$$\|P_n'\|_{C[-1,1]} \le \frac{e}{2} n \|P_n\|_{C[-1,1]} \,. \tag{4.4}$$

Earlier, Erdös [1940] proved this if all zeros of P_n are real and outside of $(-1,1)$.

The following is a useful property of the $q_{n,k}$ (see Chapter 10), proved by induction.

Proposition 4.4. *One has*

$$D^r q_{n,k}(x) = x^{k-r}(1-x)^{n-k-r} \sum_{2i+j\le r} \big[nx(1-x)\big]^i (k-nx)^j p_{i,j,r}(x)\,, \tag{4.5}$$

where $p_{i,j,r}$ are some polynomials in x with coefficients independent of n and k.

Proof. The assertion is trivial if $r=0$. In the general case, we prove (4.5) by induction. In fact, this relation implies, if Q stands for the sum above,

$$\begin{aligned}&\frac{d^{r+1}}{dx^{r+1}}\big[x^k(1-x)^{n-k}\big]\\ &\quad = x^{k-r-1}(1-x)^{n-k-r-1}\Big\{x(1-x)Q' + Q\big[(k-nx)+2rx-r\big]\Big\},\end{aligned}$$

and both terms in curly brackets have the required form

$$\sum_{2i+j\le r+1} [nx(1-x)]^i (k-nx)^j p_{i,j,r+1}(x)\,. \qquad \square$$

In the rest of this section, we present the Markov and Bernstein type inequalities for the polynomials P with positive coefficients in x, $1-x$ (Lorentz [1963]). They contain lower powers of n and of X^{-1}, where $X:=x(1-x)$, than the inequalities for unrestricted polynomials. However, the precise constants in the inequalities are not known. For the r-th derivative $P^{(r)}$, $r=1,2,\dots$ they are of the form

$$|P_n^{(r)}(x)| \le C_r n^r \|P_n\|_{C[0,1]}\,, \qquad 0\le x\le 1\,, \tag{4.6}$$

$$|P_n^{(r)}(x)| \le C_r^* X^{-r/2} n^{r/2} \|P_n\|_{C[0,1]}\,, \qquad 0<x<1\,. \tag{4.7}$$

For the sake of symmetry, it will be convenient to replace the interval $[0,1]$ by $[-1,1]$, when the $q_{n,k}$ become $\tilde q_{n,k}(x) := (1+x)^k(1-x)^{n-k}$; the polynomials with positive coefficients are then

$$\tilde P_n = \sum_{k=0}^n a_k (1+x)^k (1-x)^{n-k}\,, \qquad a_k \ge 0\,.$$

Theorem 4.5. *For the polynomials $\tilde P_n$ with positive coefficients in $(1+x)$, $(1-x)$ one has for $r=1,2,\dots$*

$$|\tilde P_n^{(r)}(x)| \le C_r \left\{\min\left(\frac{\sqrt n}{\sqrt{1-x^2}}, n\right)\right\}^r \|\tilde P_n\|_{C[-1,1]}\,, \qquad -1<x<1\,. \tag{4.8}$$

Obviously, (4.6) and (4.7) follows from this by change of variables.

The following elegant proof has been supplied by T. Erdélyi. Let

$$\Delta_{n,x} := \frac14 \max\left(\frac{\sqrt{|1-x^2|}}{\sqrt n}, \frac1n\right)\,, \qquad -1-\frac{1}{8n} \le x \le 1+\frac{1}{8n}\,.$$

Lemma 4.6. *There is an absolute constant $C>0$ so that for each polynomial $\tilde P$ of degree n with positive coefficients,*

$$|\tilde P(x+iy\Delta_{n,x})| \le C\tilde P\Big(x \pm \frac{1}{2n}\Big)\,, \quad x\in\Big[-1-\frac{1}{8n}, 1+\frac{1}{8n}\Big],\ |y|\le 1\,. \tag{4.9}$$

(Here the sign $+$ is taken if $x\le 0$, sign $-$ if $x>0$.)

Proof. We can assume that $\tilde P = \tilde q_{n,k}$ for some k, and that $0\le x\le 1+\frac{1}{8n}$. If $0\le x\le 1$, then

$$\Delta_{n,x}^2 \le \frac18\left(\frac{1-x^2}{n} + \frac{1}{n^2}\right) \le \frac{1-x}{4n} + \frac{1}{8n^2} < \frac{1}{2n}\,,$$

and this is true also if $1\le x\le 1+\frac{1}{8n}$. We obtain

$$\begin{aligned}|\tilde q_{n,k}(x+iy\Delta_{n,k})| &\le \big((1+x)^2+\Delta_{n,x}^2\big)^{k/2}\big((1-x)^2+\Delta_{n,x}^2\big)^{(n-k)/2}\\ &\le \left(1+x+\frac{1}{2n}\right)^k\left(1-x+\frac{1}{2n}\right)^{n-k}\\ &= \left(1+\left(x-\frac{1}{2n}\right)\right)^k\left(1-\left(x-\frac{1}{2n}\right)\right)^{n-k}\left(\frac{1+x+\frac{1}{2n}}{1+x-\frac{1}{2n}}\right)^k\\ &\le \tilde q_{n,k}\left(x-\frac{1}{2n}\right)\left(\frac{2n+1}{2n-1}\right)^n \le C\tilde q_{n,k}\left(x-\frac{1}{2n}\right)\end{aligned}$$

with a proper constant $C>0$. $\square$

Proof of Theorem 4.5. We denote by $\Gamma_n(x)$ the circle in $\mathbb{C}$ with center x, $-1<x<1$ and radius $\frac14\Delta_{n,x}$. Using (4.9), we obtain from Cauchy's integral formula

$$\begin{aligned}|\tilde P^{(r)}(x)| &\le \frac{r!}{2\pi}\int_{\Gamma_n(x)} \frac{|\tilde P(z)|}{|z-x|^{r+1}}|dz| \le r!\left(\frac14\Delta_{n,x}\right)^{-r} C\|\tilde P\|_{C[-1,1]}\\ &\le C(r)\left\{\min\left(\frac{\sqrt n}{\sqrt{1-x^2}}, n\right)\right\}^r \|\tilde P\|_{C[-1,1]}\,, \qquad x\in[-1,1]\,. \quad \square\end{aligned}$$

For $r=1$, the best constant in (4.6) is $C_1=e$; Scheick [1966] also finds the best $C_2=2e$.

§ 5. Lagrange Interpolation

Let $\Phi : \phi_1, \dots, \phi_n$ be a system of real-valued functions on a set A; let $x_1, \dots, x_n$ be given distinct points of A; and let $c_1, \dots, c_n$ be given real numbers. The function $P = \sum_{i=1}^n a_i \phi_i$ is said to interpolate the values of c_k if $P(x_k) = c_k$, $k = 1, \dots, n$. Usually, the c_k are values at the points x_k of some given function f; then P is said to interpolate f. Assume that we can find l_k in the linear span of Φ, $k = 1, \dots, n$ with the properties

$$l_k(x_k) = 1\,, \quad l_k(x_i) = 0\,, \quad i \neq k\,, \quad i, k = 1, 2, \dots, n\,. \tag{5.1}$$

Then P can be written in the form

$$P = \sum_{i=1}^n c_i l_i\,; \tag{5.2}$$

it is clear that $P(x_k) = c_k$ for each k. If Φ is a Haar system, P is unique. In this case, we can take

$$l_k(x) := \frac{D(x_1, \dots, x_{k-1}, x, x_{k+1}, \dots, x_n)}{D(x_1, \dots, x_{k-1}, x_k, x_{k+1}, \dots, x_n)}\,, \tag{5.3}$$

where $D(x_1, \dots, x_n)$ is the determinant (3.3) of Chapter 3. Special cases (see (5.4) and (5.6) below) may be handled by means of (5.3), or directly.

As a first example, let $x_0, \dots, x_n$ be $n+1$ distinct points of $A = [a, b]$ and let $c_0, \dots, c_n$ be given numbers. We look for an algebraic polynomial P of degree $\leq n$ that assumes values c_k at the points x_k. We can take

$$l_k(x) := \frac{(x - x_0) \cdots (x - x_{k-1})(x - x_{k+1}) \cdots (x - x_n)}{(x_k - x_0) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)}\,. \tag{5.4}$$

If we put $\Omega(x) = (x - x_0) \cdots (x - x_n)$, we can write this formula in the form

$$l_k(x) = \frac{\Omega(x)}{(x - x_k)\Omega'(x_k)}\,.$$

Thus, the interpolating polynomial is given by

$$P(x) = \sum_{k=0}^n \frac{c_k \Omega(x)}{(x - x_k)\Omega'(x_k)}\,. \tag{5.5}$$

This is Lagrange's interpolation formula; the polynomials (5.4) are its fundamental polynomials.

As another special case, let $x_0, \dots, x_{2n}$ be $2n+1$ distinct points of the circle $\mathbb{T}$. Then the trigonometric polynomial l_k of degree n, which satisfies $l_k(x_k) = 1$, $l_k(x_i) = 0$, $i \neq k$, is given by

$$l_k(x) = \frac{\sin \frac{x - x_0}{2} \cdots \sin \frac{x - x_{k-1}}{2} \sin \frac{x - x_{k+1}}{2} \cdots \sin \frac{x - x_{2n}}{2}}{\sin \frac{x_k - x_0}{2} \cdots \sin \frac{x_k - x_{k-1}}{2} \sin \frac{x_k - x_{k+1}}{2} \cdots \sin \frac{x_k - x_{2n}}{2}}\,; \tag{5.6}$$

one should notice that products of the type $\sin((x - x_0)/2)\sin((x - x_1)/2)$ can be written in the form $a + b\cos x + c \sin x$, so that l_k is indeed a polynomial of degree n.

Returning to the case of a real Haar system of n functions on $A = [a, b]$ or $A = \mathbb{T}$, we remark that *the polynomial of best approximation P of a continuous function f is also an interpolating polynomial of f at some n points of A.* In fact, by Chebyshev's theorem the difference $f - P$ takes values of alternating signs at some $n+1$ points of A; by continuity, there are at least n points where $f - P$ vanishes.

If the number of interpolation points x_k in (5.5) increases, then $n \to \infty$ and one can expect that $P_n \to f$. This is only partly true: Convergence properties of Lagrange interpolation polynomials are poor, much worse than those of Fourier series. Let

$$-1 \leq x_0^{(n)} < \cdots < x_n^{(n)} \leq 1\,, \quad n = 1, 2, \dots\,, \tag{5.7}$$

be Lagrange interpolation points in $[-1, 1]$. In 1914 Faber proved:

Theorem 5.1. *For each sequence of interpolation sets (5.7), there exists a function $f \in C[-1, 1]$ for which the interpolation polynomials $P_n(f)$ do not converge uniformly to f.*

Proof. The interpolation polynomials are linear projections of $C[-1, +1]$ onto $\mathcal{P}_n$ and the statement follows from Corollary 5.3 of Chapter 9, valid also for projections onto $\mathcal{P}_n$, and the Banach-Steinhaus theorem. □

For functions analytic on $[-1, 1]$, we can give an explicit formula for the remainder $R_n(z) := f(z) - P_n(f, z)$. Let C be a simple closed curve and let the complex interpolation points $x_k^{(n)}$ be inside C. If f is analytic on and inside C, and $\Omega_n(z) = \prod_{k=0}^n (z - x_k^{(n)})$, then

$$R_n(z) = \frac{1}{2\pi i} \int_C \frac{\Omega_n(z)}{\Omega_n(\zeta)} \frac{f(\zeta)\, d\zeta}{\zeta - z}\,. \tag{5.8}$$

This follows from the calculus of residues. The residue of the integrand at its simple pole $\zeta = z$ is $f(z)$, those at the $x_k^{(n)}$ are equal to

$$f(x_k^{(n)})\Omega_n(z)/(x_k^{(n)} - z)\Omega_n'(x_k^{(n)})\,,$$

and we obtain (5.8) via (5.5).

If the interpolation points and the curve C match each other properly, one can obtain convergence $P_n(z) \to f(z)$ in the whole region bounded by C. This is the case for the Chebyshev interpolation points, that is, zeros of the Chebyshev polynomials C_n given in (6.4) of Chapter 3 and the ellipses E_ρ of Section 2. For those points, $\Omega_n(z) = 2^{-n+1} C_n(z)$. From (6.2) of Chapter 3, we have $C_n(z) = (w^n + w^{-n})/2$ whenever $z = (w + w^{-1})/2$. Hence $\lim_{n\to\infty} \sqrt[n]{|\Omega_n(z)|} = \frac{1}{2}\rho$ for $z \in E_\rho$, $\rho > 1$.

Theorem 5.2. *If f is analytic on the closed ellipse E_ρ, $\rho > 1$, and* (5.7) *are the Chebyshev interpolation points, then $P_n(f,z) \to f(z)$ uniformly for $z \in E_\rho$.*

Proof. There is an ellipse $E_{\rho_1} \supset E_\rho$, $\rho_1 > \rho$, on which f is analytic. Let C be the boundary of E_{ρ_1}, let δ be the distance from C to E_ρ, then for all $\varepsilon > 0$ and all large n, $|\Omega_n(z)| \le (\frac{\rho+\varepsilon}{2})^n$, $z \in E_\rho$ and $|\Omega_n(\zeta)| \ge (\frac{\rho_1-\varepsilon}{2})^n$ for $\zeta \in C$, and we can estimate the integral in (5.8): $|R_n(z)| \le \mathrm{Const}\,\frac{1}{\delta} q^n \to 0$, where $q = \frac{\rho+\varepsilon}{\rho_1-\varepsilon} < 1$. □

The Chebyshev interpolation points are from many points of view "the best" interpolation points on $[-1,1]$. For these points, Grünwald and Marcienkewicz proved in 1937 independently the existence of a function $f \in C(-1,1)$ for which the interpolation polynomials diverge everywhere on $[-1,1]$.

This fact strongly suggests that the same situation prevails for any set of interpolation points (5.7). The proof, which presents great technical difficulties was found only in 1980 by Erdös and Vertesi [1981].

Theorem 5.3. *For each sequence of interpolation sets* (5.7), *there is a function $f \in C[-1,1]$ whose Lagrange interpolation polynomials diverge almost everywhere on $[-1,1]$.*

The proof cannot be given here. One cannot assert divergence everywhere (instead of almost everywhere) because some $c \in [-1,1]$ may appear in the set (5.7) for each large n, and then $P_n(f,c) = f(c)$ for such n.

In contrast to uniform convergence, convergence $P_n \to f$ in the L_2-norm can be proved in the important special case when the interpolation points (5.7) are zeros of the $n+1$-st orthogonal polynomial with respect to some non-negative weight $w \in L_1(-1,1)$. (What we need to know about the zeros of orthogonal polynomials is only that they are real and lie in $[-1,1]$.) In this case Ω_n is a constant multiple of the $n+1$-st orthogonal polynomial. The fundamental interpolation polynomials l_k of (5.5) are then orthogonal to each other: for $j \ne k$,

$$\int_{-1}^{+1} w l_j l_k\,dx = \mathrm{Const}\int_{-1}^{1} w(x)\Omega_n(x)\frac{\Omega_n(x)}{(x-x_j)(x-x_k)}\,dx = 0\,, \tag{5.9}$$

because Ω_n is orthogonal on $[-1,1]$ to each polynomial of degree $< n$. Another useful formula

$$\int_{-1}^{+1} w \sum_{k=0}^{n} l_k^2\,dx = \|w\|_1 \tag{5.10}$$

follows by squaring and then integrating the relation $\sum_0^n l_k(x) = 1$.

Theorem 5.4 (Erdös and Turán [1937]). *For each $f \in C[-1,1]$,*

$$\int_{-1}^{1} w\big(f - P_n(f)\big)^2\,dx \to 0\,, \quad n \to \infty\,. \tag{5.11}$$

Proof. Let Q be the polynomial of degree $\le n$ of best uniform approximation for f, so that $|f(x) - Q(x)| \le E_n$, $x \in [-1,1]$. Then

$$\int_{-1}^{1} w|f - Q|^2\,dx \le E_n^2\|w\|_1\,. \tag{5.12}$$

Since $Q = P_n(Q)$, we estimate by means of (5.9) and (5.10) the integral

$$\begin{aligned}\int_{-1}^{1} w|P_n(f) - P_n(Q)|^2\,dx &= \int_{-1}^{1} w(x)\Big\{\sum_{k=0}^{n}(f(x_k) - Q(x_k))l_k(x)\Big\}^2\,dx\\ &= \int_{-1}^{1} w\sum_{k=0}^{n}\big(f(x_k) - Q(x_k)\big)^2 l_k^2\,dx\\ &\le E_n^2\int_{-1}^{1} w\sum l_k^2\,dx = E_n^2\|w\|_1\,,\end{aligned}$$

hence

$$\int_{-1}^{1} w\big(f - P_n(f)\big)^2\,dx \le 4E_n^2\|w\|_1 \to 0\,. \qquad \square$$

If, in particular, w is bounded away from zero in $[-1,1]$, for instance if $w(x) = (1-x^2)^{-1/2}$, we obtain

$$\int_{-1}^{1}\big(f - P_n(f)\big)^2\,dx \to 0\,.$$

The convergence $P_n(f) \to f$ in spaces L_p with $p > 2$ can also sometimes be established, as has been done by Askey and by Nevai, see Nevai [1984]. But this is decidedly more difficult.

§ 6. Hermite Interpolation

In this interpolation, not only the values of the polynomial $P_n(x) = a_0 + \cdots + a_n x^n/n!$, but also the values of some of its successive derivatives are prescribed. Let $y_1, \ldots, y_p$ be distinct real or complex *interpolation points*, which are equipped with multiplicities $m_j > 0$, $j = 1, \ldots, p$, with $m_1 + \cdots + m_p = n+1$. Let $c_{j,l}$ be given constants. We look for a P_n which satisfies the equations

$$\text{(6.1)} \qquad \begin{cases} P_n^{(l)}(y_j) = a_l + a_{l+1}y_j + \cdots + a_n y_j^{n-l}/(n-l)! = c_{j,l}\,, \\ l = 0,\dots,m_j - 1\,, \quad j = 1,\dots,p\,. \end{cases}$$

In particular, for some function f we may take $c_{j,l} = f^{(l)}(y_j)$. Then the polynomial $P_n := P_n(f,x)$ *interpolates the function* f.

Theorem 6.1. *There is a unique polynomial P_n which solves* (6.1).

Proof. We view (6.1) as a system of equations for the unknown coefficients a_k. If all the $c_{j,l}$ are zero, then P_n has $(n+1)$ zeros, and hence is zero. This shows that the determinant of the system (6.1) is different from zero. Therefore, there exists a unique solution P_n for any choice of the $c_{j,l}$. □

Sometimes a slightly different point of view is preferable. We will adhere to it in this and the next section. Let $X : x_0,\dots,x_n$ be the interpolation points, with possible repetitions. For each j, the multiplicity m_j of x_j is the number of $x_i = x_j$, while l_j is the number of $x_i = x_j$ with $i \le j$. Then equations (6.1) are replaced by

$$\text{(6.2)} \qquad P_n^{(l_j-1)}(x_j) = c_j\,, \quad j = 0,\dots,n$$

with properly chosen c_j.

The interpolation polynomial $P_n(x) := P_n(f,X)(x) := P_n(f,X;x)$ depends on the $(n+1)$-tuple X. It can be obtained by means of Newton's method:

Theorem 6.2. *There exist unique constants $A_0,\dots,A_n$ for which the polynomials*

$$\text{(6.3)} \qquad \begin{cases} P_0(x) := A_0 \\ P_1(x) := A_0 + A_1(x - x_0) \\ \cdots\cdots\cdots \\ P_n(x) := A_0 + A_1(x-x_0) + \cdots + A_n(x-x_0)\cdots(x-x_{n-1}) \end{cases}$$

are the solutions of the Hermite interpolation problems for the sets of interpolation points $X_0 := \{x_0\},\dots,X_n := \{x_0,\dots,x_n\}$ and given data $c_0,\dots,c_n$.

Proof. We proceed by induction on k. Certainly $P_0(x) := A_0$ is uniquely defined by the condition $P_0(x_0) = c_0$.

Suppose that our assertion is true for $P_0,\dots,P_{k-1}$ and $X_0,\dots,X_{k-1}$. The additional point x_k in X_k is associated with the condition $P_k^{(l_k-1)}(x_k) = c_k$ for the polynomial $P_k(x) = P_{k-1}(x) + A_k(x-x_0)\cdots(x-x_{k-1})$. This defines A_k uniquely; from Leibniz's formula it follows that the (l_k-1)-st derivative of $(x-x_0)\cdots(x-x_{k-1})$ is non-zero at x_k. The other conditions (6.2) for P_k and $i < k$ are satisfied because they are satisfied for P_{k-1} and because the $(l_i - 1)$-st derivative of $(x-x_0)\cdots(x-x_{k-1})$ is zero at x_i. □

Extreme cases of Hermite interpolation are Lagrange interpolation, when all x_i are different and Taylor interpolation, when $x_0 = \cdots = x_n$. In the latter case, if $c_k = f^{(k)}(x_0)$, $k = 0,\dots,n$, P_n is the Taylor polynomial

$$P_n(x) = f(x_0) + f'(x_0)(x-x_0) + \cdots + f^{(n)}(x_0)(x-x_0)^n/n!$$

Theorem 6.1 implies that the determinant of the system (6.2)

$$\text{(6.4)} \qquad D(X) := \det\left[0,\dots,0,1,\dots,x_j^{n-l_j+1}/(n-l_j+1)!\right]_{j=0}^{n}$$

is different from zero. Solving (6.2) with respect to a_k we get

$$\text{(6.5)} \qquad a_k = \frac{D_k(X)}{D(X)}$$

where $D_k(X)$ is obtained from $D(X)$ by replacing its k-th column by the column of the c_i. In the case that $c_i = f^{(l_i-1)}(x_i)$ we see that the a_k *are linear combinations of these derivatives with coefficients that are rational functions of the x_i's.*

We shall study continuity (in the uniform norm) of the interpolating polynomial $P_n(f,X)$ for a function $f \in C[a,b]$, as a function of $X = (x_0,\dots,x_n)$. We assume that the x_i are real with $a \le x_i \le b$, then X is a point of $[a,b]^{n+1}$. We measure the distance between two points X, X' by their Euclidean distance. Each $(n+1)$-tuple X defines its own multiplicities m_j, numbers l_j, $j = 0,\dots,n$, and its own set of equations (6.2) with $c_i = f^{(l_i-1)}(x_i)$.

Theorem 6.3. *For a given $f \in C[a,b]$, the polynomial $P_n(f,X)$ is a continuous function of X at each point $X = (x_0,\dots,x_n)$ that has a neighborhood in which the derivatives $f^{(m_j-1)}$ exist and are continuous in the uniform norm on $[a,b]$.*

Proof (Dyn, Lorentz and Riemenschneider [1982]). For an $(n+1)$-tuple X of this type, we take $\delta > 0$ so small that for $Y = (y_0,\dots,y_n) \in [a,b]^{n+1}$ with $\|Y - X\| < \delta$ all derivatives $f^{(m_j-1)}(y_j)$ exist and are continuous. For these Y we keep the form of the equations (6.2) corresponding to X and define the polynomial $Q_n(x) = \sum_{k=0}^n b_k x^k/k!$ by means of the equations

$$\text{(6.6)} \qquad Q_n^{(l_j-1)}(y_j) = f^{(l_j-1)}(y_j)\,, \quad j = 0,\dots,n\,,$$

(with the l_j defined by X). The determinant of the equations (6.6) is defined and is close to $D(X)$ provided $\delta > 0$ is sufficiently small. Since $D(X) \ne 0$, these equations can be solved for the b_k; the resulting b_k are close to the coefficients a_k of $P_n(f,X)$. As a consequence, $\|Q_n - P_n\| < \varepsilon$ for $\|Y - X\| < \delta$.

Now let $X' = (x'_0,\dots,x'_n)$ be another set of interpolation points for which $\|X - X'\| < \delta$. If the multiplicity of an x_i is m_i and $\delta > 0$ is small, there are precisely m_i points x'_j for which $|x_i - x'_j| < \delta$. The definition of the polynomial $P_n(f,X')$ shows that $g(x) = P_n(f,X';x) - f(x)$ has (counting multiplicity) at least m_i zeros satisfying $|x_i - x| < \delta$. By Rolle's theorem, and the continuity of the derivatives of f, there are points z_l in this interval for which $g^{(l)}(z_l) = 0$, $l = 0, \dots, m_i - 1$. We see that $P_n(f, X')$ is one of the polynomials Q_n and consequently $\|P(f,X) - P(f,X')\| < \varepsilon$. □

§ 7. Divided Differences

For a function f and points $X : x_0, \dots, x_n$ (not necessarily ordered by $\le$), we define the *n-th divided difference of f* by

$$[x_0, \dots, x_n]f := A_n \tag{7.1}$$

where A_n is the coefficient of x^n of the polynomial $P_n(f, X, x)$ which interpolates f at $x_0, \dots, x_n$; A_n is given by (6.3). This definition requires that all derivatives $f^{(l)}(x_i)$ which appear in (6.2) be defined; $[x_0, \dots, x_n]f$ is a linear combination of these derivatives. For example $[x_0]f = f(x_0)$, $[x_0, x_1]f = \frac{f(x_1) - f(x_0)}{x_1 - x_0}$ if $x_1 \ne x_0$ and $[x_0, x_0]f = f'(x_0)$ if this derivative exists.

Many properties of divided differences follow from those of the interpolating polynomials of § 6. The divided difference $[x_0, \dots, x_n]f$ is a linear combination of the derivatives $f^{(k)}(x_i)$, $0 \le k \le m_i - 1$, where m_i is the multiplicity of the point x_i, $i = 0, \dots, n$. They have the properties:

(a) $[x_0, \dots, x_n]f$ is symmetric in $x_0, \dots, x_n$.

(b) $[x_0, \dots, x_n]f$ is a constant if f is a polynomial of degree $\le n$, and is zero for a polynomial of degree $< n$.

(c) $[x_0, \dots, x_0]f = f^{(n)}(x_0)/n!$

(d) The coefficients A_k in (6.3) are the divided differences $[x_0, \dots, x_k]f$. This leads to *Newton's formula* for the interpolating polynomial:

$$P_n(f, X; x) = \sum_{k=0}^{n} (x - x_0) \cdots (x - x_{k-1})[x_0, \dots, x_k]f\,. \tag{7.2}$$

If we adjoin an additional point y to X and let $Y : x_0, \dots, x_n, y$, we will have $P_{n+1}(f, Y; y) = f(y)$. We obtain from this a formula with remainder

$$f(y) = P_n(f, X; y) + (y - x_0) \cdots (y - x_n)[x_0, \dots, x_n, y]f\,. \tag{7.3}$$

(e) If $f \in C^n[a,b]$, $a \le x_i \le b$, $i = 0, \dots, n$, then by Rolle's theorem

$$[x_0, \dots, x_n]f = f^{(n)}(\xi)/n! \quad \text{for some} \quad a \le \xi \le b\,. \tag{7.4}$$

(f) $[x_0, \dots, x_n]f$ is continuous at the point $(x_0, \dots, x_n)$, if the derivatives of f of proper orders are continuous at the x_i.

(g) For $x_0 \ne x_n$, we have the recurrence relation

$$[x_0, \dots, x_n]f = \frac{1}{x_n - x_0}\{[x_1, \dots, x_n]f - [x_0, \dots, x_{n-1}]f\}\,. \tag{7.5}$$

This explains the name "divided difference". With formulas (c) and (7.5) all differences can be computed. For the proof of (7.5), let S, T be polynomials of degrees $\le n-1$, which interpolate f at the points $x_0, \dots, x_{n-1}$ and $x_1, \dots, x_n$, respectively. Then

$$\frac{x - x_0}{x_n - x_0} T(x) + \frac{x_n - x}{x_n - x_0} S(x) \tag{7.6}$$

is a polynomial of degree $\le n$, and an application of Leibniz's formula shows that it interpolates f at $x_0, \dots, x_n$ with the required multiplicities. Comparing the leading terms of $P_n(f, X)$ and of (7.6), we obtain (7.5).

(h) If the x_i are all different,

$$\begin{aligned}[x_0, \dots, x_n]f &= \sum_{k=0}^{n} \frac{f(x_k)}{(x_k - x_0) \cdots (x_k - x_{k-1})(x_k - x_{k+1}) \cdots (x_k - x_n)} \\ &= \sum_{k=0}^{n} \frac{f(x_k)}{\Omega'(x_k)}\end{aligned} \tag{7.7}$$

with Ω from (5.5). One obtains this from (5.5) with $c_k := f(x_k)$.

(i) The *differences* of a function f with step h are defined by $\Delta_h^0 f(x) := f(x)$, $\Delta_h f(x) := \Delta_h^1 f(x) := f(x+h) - f(x)$, and $\Delta_h^n f(x) := \Delta_h \Delta_h^{n-1} f(x)$. Then

$$\Delta_h^n f(x) = \sum_{k=0}^{n} (-1)^{n-k} \binom{n}{k} f(x + kh)\,. \tag{7.8}$$

This can be proved directly or deduced from (7.7) if one notices that

$$\Delta_h^n f(x) = n!\, h^n [x, x+h, \dots, x+nh]f\,. \tag{7.9}$$

(j) The following is the *Leibniz formula for divided differences*:

$$[x_0, \dots, x_n](gh) = \sum_{k=0}^{n} \big([x_0, \dots, x_k]g\big)\big([x_k, \dots, x_n]h\big)\,. \tag{7.10}$$

For the proof, we first assume that the x_i are distinct, and consider the polynomial

$$\begin{aligned} P_n(g, X; x) P_n(h, X; x) &= \sum_{k=0}^{n} (x - x_0) \cdots (x - x_{k-1})[x_0, \dots, x_k]g \\ &\times \sum_{l=0}^{n} (x - x_{l+1}) \cdots (x - x_n)[x_l, \dots, x_n]h \end{aligned} \tag{7.11}$$

which interpolates $f = gh$ at the points x_i. We write the product of the two sums as

$$\sum_{k,l=0}^{n} = \sum_{k \le l} + \sum_{k > l}\,.$$

The second of the sums vanishes at $x = x_0, \dots, x_n$. Hence the first sum, which is a polynomial Q of degree $\le n$, also interpolates f at the x_i. By unicity, Q is $P_n(f, X; x)$. Comparing the leading coefficients in $P_n(f, X; x)$ and Q we obtain (7.10). If some of the x_i coincide, we *assume that the derivatives of g and h, required for the divided difference $[x_0, \dots, x_n]$, are continuous at the corresponding points*. Approximating the set $X : x_0, \dots, x_n$ by a set $X' : x_0', \dots, x_n'$ with distinct x_i', we obtain (7.10) by continuity (Theorem 6.3).

(k) Another useful formula is the integral representation

$$(7.12)\qquad \begin{aligned}[x_0, \dots, x_n]f = \int_0^1 dt_1 \int_0^{t_1} dt_2 \cdots \int_0^{t_{n-1}} f^{(n)}\big(x_0 + (x_1 - x_0)t_1 + \cdots \\ + (x_n - x_{n-1})t_n\big)\, dt_n\end{aligned}$$

valid if $f^{(n-1)}$ is absolutely continuous, and if not all x_i coincide.

Indeed, this can be proved by induction on n. First, let the x_i be distinct, then

$$\int_0^1 f'\big(x_0 + (x_1 - x_0)t_1\big)\, dt_1 = \frac{1}{x_1 - x_0} \int_{x_0}^{x_1} f'(u)\, du = [x_0, x_1]f .$$

Assume that (7.12) holds with n replaced by $n-1$. After one integration with $u_0 = x_0 + \cdots + (x_{n-1} - x_{n-2})t_{n-1}$, $u_1 = x_0 + \cdots + (x_{n-2} - x_{n-3})t_{n-2} + (x_n - x_{n-2})t_{n-1}$ the right-hand side of (7.12) becomes

$$\begin{aligned}&\int_0^1 dt_1 \cdots \int_0^{t_{n-2}} \frac{f^{(n-1)}(u_1) - f^{(n-1)}(u_0)}{x_n - x_{n-1}}\, dt_{n-1} \\ &= \frac{1}{x_n - x_{n-1}} \big\{[x_0, \dots, x_{n-2}, x_n]f - [x_0, \dots, x_{n-2}, x_{n-1}]f\big\} = [x_0, \dots, x_n]f .\end{aligned}$$

As long as not all x_i coincide, both sides of (7.12) are continuous functions, so that the formula is valid in general.

(l) If we apply (7.12) to the function $f(x) = x^n/n!$, we obtain that the integral of (7.12), with $f^{(n)}$ replaced by 1, is equal to $1/n!$. Therefore, under the assumptions of (k)

$$(7.13)\qquad |[x_0, \dots, x_n]f| \le \frac{1}{n!} \|f^{(n)}\|_\infty .$$

If $\lambda(f) = \int_a^b f\, d\mu$ is a continuous linear functional on the space $C[a, b]$ which is orthogonal to all polynomials P_{n-1} of degree $\le n-1$: $\lambda(P_{n-1}) = 0$, then for $f \in C^n[a, b]$,

$$(7.14)\qquad \lambda(f) = \int_a^b f^{(n)}(t) \lambda \left[\frac{(\cdot - t)_+^{n-1}}{(n-1)!}\right] dt .$$

This can be obtained by applying λ to both sides of the Taylor formula

$$(7.15)\qquad \begin{aligned}f(x) = f(a) + \frac{x - a}{1!} f(a) + \cdots + \frac{(x-a)^{n-1}}{(n-1)!} f^{(n-1)}(a) \\ + \int_a^b f^{(n)}(t) \frac{(x - t)_+^{n-1}}{(n-1)!}\, dt ,\end{aligned}$$

because the integrals with respect to dt and $d\mu(x)$ are interchangeable.

In a representation of the type

$$\lambda(f) = \int_a^b f^{(n)}(t) K(t)\, dt ,$$

the kernel K is called a Peano kernel. We can find the Peano kernel of the divided differences.

If all the points $x_0, \dots, x_n$ are distinct, the functional $\lambda(f) = [x_0, \dots, x_n]f$ is continuous on $C[a, b]$. Hence, from (7.14),

$$(7.16)\qquad [x_0, \dots, x_n]f = \int_a^b f^{(n)}(t) [x_0, \dots, x_n] \left[\frac{(\cdot - t)_+^{n-1}}{(n-1)!}\right] dt .$$

From (f), both sides of (7.16) are continuous provided not all of the x_i coincide. Hence (7.16) is valid in this case as well.

§ 8. Quadrature Formulas

A *quadrature formula* is an expression of the form

$$(8.1)\qquad I(f) = \sum_{k=1}^n c_k f(x_k) ,$$

if it is treated as an approximation to the integral $\int_A f\, d\mu$. Here x_k, $k = 1, \dots, n$, are points in A, μ is a measure on A. Quadrature formulas are related to interpolation formulas; some of them can be obtained by integrating the latter.

In particular, we apply this idea to the Lagrange interpolation (5.5) by algebraic polynomials P on $[a, b]$. Let $\Omega(x) = (x - x_1) \cdots (x - x_n)$. The polynomials $l_k(x) = \Omega(x_k)/[(x - x_k)\Omega'(x_k)]$, $k = 1, \dots, n$, which appear in (5.5) are a basis for $\mathcal{P}_n$. Therefore, *formula* (8.1) *is exact for all P of degree $\le n$, that is, we have*

$$(8.2)\qquad I(P) = \int_a^b P\, d\mu , \quad \textit{for all} \quad P \in \mathcal{P}_n$$

if and only if (8.2) *is valid for each l_k, in other words, if and only if*

$$(8.3)\qquad c_k = \int_a^b \frac{\Omega(x)}{(x - x_k)\Omega'(x)}\, d\mu(x) , \quad k = 1, \dots, n .$$

Formulas (8.1) of this type are called *interpolating quadrature formulas*.

The n points x_k in (8.1) have been arbitrary. By selecting them properly, one can increase the dimension of the space of polynomials, for which (8.2) is valid. Formula (8.1) is called a *Gaussian formula*, if it is exact for all polynomials of degree $\le 2n-1$.

Theorem 8.1. *Formula* (8.1) *is a Gaussian formula with respect to* $d\mu$ *on* $[a,b]$ *if and only if it is interpolatory and* Ω *is orthogonal with respect to* $d\mu$ *to all polynomials of degree* $< n$.

Proof. (a) *Necessity.* It is necessary that (8.1) should be an interpolatory formula. Now let $P = Q\Omega$, where Q is a polynomial of degree $< n$. Then $P(x_k) = 0$, and so $I(P) = 0$. We must have $\int_a^b Q\Omega\,d\mu = 0$.

(b) *Sufficiency.* We can write an arbitrary polynomial P of degree $\le 2n-1$ as $P = Q\Omega + R$, where Q, R are of degree $< n$. Then $P(x_k) = R(x_k)$, $k = 1,\dots,n$, and thus $I(P) = I(R) = \int_a^b R\,d\mu = \int_a^b P\,d\mu$, since $\int_a^b Q\Omega\,d\mu = 0$. □

Gaussian formulas are related to the theory of orthogonal polynomials. To each positive measure $d\mu$ on $[a,b]$ there corresponds a sequence of polynomials P_n, $n = 0,1,2,\dots$, of degree n, which is orthonormal with respect to the measure $d\mu$; then $\int_a^b P_k P_l\,d\mu = \delta_{k,l}$, $k,l = 0,1,\dots$. For this theory, compare the books Szegö [B-1975] and Freud [B-1969]. One of the first results of this theory is that the zeros of each P_n are real, simple, and lie in (a,b). By what we have proved, (8.1) is then a Gaussian quadrature formula. To the measure $d\mu = (1-x^2)^{-1/2}\,dx$ on $[-1,1]$ correspond the Chebyshev polynomials (see §6 in Chapter 3); to the Lebesgue measure dx – the Legendre polynomials. To each of these systems corresponds its own Gaussian formula.

§ 9. Birkhoff Interpolation

In Hermite interpolation, we prescribe consecutive derivatives at interpolation points. Polynomial interpolation with the possibility of gaps for the orders of derivatives has been studied for the first time by G.D. Birkhoff [1906] in 1906, but it remained dormant for some some 60 years. Birkhoff himself was mainly interested in the remainder formulas for this interpolation, for example the "Peano interpolation kernels" appear for the first time in his work.

Schoenberg [1966] has formalized the problem as follows. Let $E = (e_{i,k})_{i=1}^{m}{}_{k=0}^{n}$ be an $m\times(n+1)$ *interpolation matrix* with elements $e_{i,k}$ that are zeros or ones, with exactly $n+1$ ones. Let $X = (x_1,\dots,x_m)$, $x_1 < \cdots < x_m$ be a set of m distinct *interpolation points*. For polynomials $P(x) = \sum_{k=0}^n a_k x^k/k!$ of degree $\le n$ we consider the $n+1$ interpolation equations

$$P^{(k)}(x_i) = c_{i,k} \quad \text{if} \quad e_{i,k} = 1 \tag{9.1}$$

(the $c_{i,k}$ are given data) and seek to determine the unknown coefficients a_k. In contrast to Hermite interpolation, this is not always possible.

The matrix of coefficients M of (9.1) and its determinant D are given by

$$\begin{cases} D := D(E,X) := \det M(E,X)\,, \\ M := M(E,X) := \left[x_i^{-k}/(-k)!,\dots,x_i^{n-k}/(n-k)!;\, e_{i,k}=1\right]. \end{cases} \tag{9.2}$$

We agree to interpret terms containing $1/r!$ to be zero if $r < 0$. Formula (9.2) gives the row of M which corresponds to a pair (i,k) with $e_{i,k} = 1$; there are $n+1$ rows, we order them lexicographically with respect to i,k. The columns of M give the derivatives of the terms $x^l/l!$, $l = 0,\dots,n$, of P, evaluated at the x_i. The *pair* E, X *is regular, if* (9.1) *has a solution for all values of* $c_{i,k}$. This is so if and only if $D(E,X) \ne 0$ for the set of interpolation points X. The *matrix* E *is regular if* E, X *is regular for each set* X.

There is a related useful property of the matrix E. As a function of $x_1,\dots,x_m$, $D(E,X)$ is a polynomial (of high degree) of the x_i. Either this polynomial is identically zero on $\mathbf{R}^m$, or it is zero only on a set of m-dimensional measure zero, which is also of first category in $\mathbf{R}^m$. In this case we call E *almost regular*: then one can hope to get a regular pair E, X by selecting the points X at random.

Examples. (a) The following matrices are interpolation matrices for polynomials of degree ≤ 2:

$$E_1 = \begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix}, \quad E_2 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & 0 \end{pmatrix}, \quad E_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix},$$

for $X_1 = (x_1,x_2)$, $X_2 = X_3 = (x_1,x_2,x_3)$. Here, E_1 is a regular Hermite matrix; for E_2 we have $D(E_2,X) \equiv 0$; for E_3, $D(E_3,X) = \frac{1}{2}(x_3-x_1)(x_3+x_1-2x_2)$ if $X = (x_1,x_2,x_3)$. Thus, E_3 is almost regular; the pair E_3, X is regular if and only if $x_2 \ne \frac{1}{2}(x_1+x_3)$.

(b) The $1\times(n+1)$ matrix with all elements $= 1$ is the *regular Taylor matrix*. Also, the $(n+1)\times 1$ *Lagrange interpolation matrix* with elements one is regular.

(c) The matrix

$$E_4 = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 \end{pmatrix} \tag{9.3}$$

for polynomials of degree 5 is regular, since with $X = (0,x,1)$,

$$D(E_4,X) = \text{Const}\,x(x-1)(15x^2-15x+4) \ne 0, \quad 0 < x < 1\,.$$

(An arbitrary set X can be reduced to the case $x_1 = 0$, $x_m = 1$ by a linear transformation.)

Let

$$M_r := \sum_{k=0}^{r} \sum_{i=1}^{m} e_{i,k}\,, \quad r = 0, \ldots, n\,, \tag{9.4}$$

be the number of ones in the first $r+1$ columns of E. (Thus, $M_n = n+1$.) The conditions

$$M_r \geq r+1\,, \quad r = 0, \ldots, n\,, \tag{9.5}$$

are called the *Pólya conditions*, and any matrix E satisfying (9.5) is a *Pólya matrix.* The conditions (9.5) are necessary for E to be almost regular. Indeed, if $M_r \leq r$, then the first $r+1$ columns of $M(E,X)$ are linearly dependent and therefore $D(E,X) = 0$ for each X.

If we subtract both sides of (9.5) from $n+1$, we obtain another, *equivalent form of the Pólya condition*:

$$\mu_r \leq n - r\,, \quad r = 0, \ldots, n\,, \tag{9.6}$$

where μ_r stands for the number of ones in columns $r+1, \ldots, n$ of E.

Some operations performed on a Pólya matrix E preserve this property. To describe them, we introduce the following notion. A *shift* Λ of E, and of the i-th row of E is a translation of a one, $e_{i,r} = 1$, of this row one step to the right, to the place $(i, r+1)$. This assumes that $r+1 \leq n$ and that the place $(i, r+1)$ is not occupied, that is, that $e_{i,r+1} = 0$. The shift produces a new matrix ΛE.

We shall say that two rows of E, for example, the first and the second, *have a collision*, if they have ones in the same column: $e_{1,k} = e_{2,k} = 1$. *In this case there is a shift of row* 1 *which preserves the Pólya property of* E. Indeed, let $s \geq 0$ be defined by $1 = e_{1,k} = \cdots = e_{1,k+s}$, $e_{1,k+s+1} = 0$. This s exists, for otherwise the Pólya condition (9.6) would be violated. Then $M_r > r+1$ for $r = k, \ldots, k+s$. The shift Λ of $e_{1,k+s} = 1$ will decrease M_{k+s} by one, will increase M_{k+s+1} by one, and will not change the other values M_r. It will preserve (9.5).

Let E_1, E_2 be matrices consisting of the first $r+1$ and the last $n-r$ columns of E, respectively. We call $E = E_1 \oplus E_2$ a *decomposition* of E, if each of the submatrices E_1, E_2 has as many ones as it has columns.

Theorem 9.1 (Atkinson-Sharma [1969], Ferguson [1969]). *If $E = E_1 \oplus E_2$ is a decomposition of E, then E is regular, or almost regular whenever both E_1 and E_2 have this property.*

Proof. Interchanging rows of $D(E,X)$, we can make the first $r+1$ rows of the determinant correspond to ones in the matrix E_1, the last $n-r$ rows to ones in E_2, with the corresponding subsets X_1, X_2 of X. Then $M(E,X)$ will decompose

$$M(E,X) = \begin{array}{c} \left(\begin{array}{cc} M(E_1,X_1) & * \\ O & M(E_2,X_2) \end{array} \right) \begin{array}{l} \} \; r+1 \\ \} \; n-r \end{array} \\ \begin{array}{cc} \underbrace{\qquad\quad}_{r+1} & \underbrace{\qquad\quad}_{n-r} \end{array} \end{array}$$

It follows that

$$D(E,X) = \pm D(E_1,X_1) D(E_2,X_2)\,. \tag{9.7}$$

This yields both statements of the theorem. □

Example. An *Abel matrix* E contains a single one in each of its columns. This means that each derivative $P^{(k)}$, $k = 0, \ldots, n$ appears exactly once in (9.1). In this case, E decomposes into $n+1$ one column Lagrange matrices, each containing a one. Hence E is regular.

In the next section we will encounter many other examples of regular matrices.

§ 10. Regularity of Birkhoff Matrices

The following Theorem 10.1 which characterizes almost regular matrices was given by Birkhoff [1906] but without a valid proof. Our proof is based on the differentiation of the determinant $D(E,X)$. First we shall see how the derivatives of $D(E,X)$ are related to the shifts of § 9. We fix the interpolation points $x_2, \ldots, x_m$ and differentiate D with respect to x_1. Now D may have several rows containing x_1 (and corresponding to a single row of E); $\frac{dD}{dx_1}$ is a sum of several determinants, with one of the rows differentiated in each. If we differentiate the row of D with $e_{1,k} = 1$, and if $k = n$, we obtain a row that is identically zero. Similarly, if $e_{1,k+1} = 1$, we get a determinant with two equal rows which is zero. If $e_{1,k+1} = 0$, this differentiation produces $D(\Lambda E, X)$, with the *shift* Λ of row 1 in E taking $e_{1,k} = 1$ into the new position $(1, k+1)$. Finally, if ΛE is not a Pólya matrix, the determinant $D(\Lambda, E, X)$ is identically zero by § 9. We see that $\frac{d}{dx_1} D = \sum D(\Lambda E, X)$, where the sum is extended over all admissible shifts Λ of row 1 in E, for which ΛE is a Pólya matrix.

More generally,

$$\frac{d^\alpha}{dx_1^\alpha} D = \sum D(\Lambda^\alpha E, X) = \sum D(\Lambda_\alpha \ldots \Lambda_1 E, X) \tag{10.1}$$

for all shifts Λ^α of row 1 in E. Here the *multiple shift* Λ^α of order α is a product of α simple shifts, $\Lambda^\alpha = \Lambda_\alpha \ldots \Lambda_1$; it is assumed that the shifts Λ_1 of E, Λ_2 of $\Lambda_1 E, \ldots$ are all defined. Two Λ^α are equal if the corresponding single shifts are pairwise equal. The sum in (10.1) extends over all possible shifts of order α, for which $\Lambda^\alpha E$ is a Pólya matrix.

The effect of Λ^α on the row 1 of E is as follows. If $(1,k_1),\dots,(1,k_p)$ are the original positions of ones in this row, Λ^α will move them to the new positions $(1,l_1),\dots,(1,l_p)$ with $k_i \le l_i$ and

$$\alpha = \sum_{i=1}^{p}(l_i - k_i)\,. \tag{10.2}$$

The following theorem (with a different proof) is due to Nemeth [1966].

Theorem 10.1. *An interpolation matrix E is almost regular if and only if it satisfies the Pólya condition.*

Proof. We have to show only the sufficiency of the condition. We achieve this by obtaining a formula for $d^\alpha D(E,X)/dx_1^\alpha$ for a specific X and using induction on m. If $m=1$, the statement is obvious, for then E is a Taylor matrix and $D(E,X)\equiv 1$.

In the general case, we use (10.1), for the derivative

$$\frac{d^\alpha}{dx_1^\alpha}D(E,X)_{x_1=x_2} = \sum D(\Lambda^\alpha E, X)_{x_1=x_2}\,. \tag{10.2}$$

In the sum (10.2), we need to take into account only those Λ^α for which rows 1 and 2 of $\Lambda^\alpha E$ have no collision, for otherwise the substitution $x_1 = x_2$ will make $D(\Lambda^\alpha E, X)$ equal to zero. We shall show that it is possible to find α so that all nonzero terms of the sum are equal to $\varepsilon D(E^*,X^*)$, with the same $\varepsilon = \pm 1$ and with an $(m-1)\times(n+1)$ Pólya matrix E^*.

If $(1,k_i)$, $i=1,\dots,p$, are the positions of all of the ones in row 1, we define $0 \le l_1^* < \dots < l_p^* \le n$ as follows. Let l_1^* be the first $l \ge k_1$ with $e_{2,l}=0$. The remaining l_i^* are defined by induction: for $i>1$, l_i^* is the smallest l for which $l > l_{i-1}^*$, $l \ge k_i$ and $e_{2,l}=0$. All l_i^*, $i=1,\dots,p$, exist, for otherwise condition (9.6) would be violated. We let $\alpha := \sum_{i=1}^p (l_i^* - k_i)$.

Let E^* be the matrix E with row one omitted and with ones $e_{1,k_i}=1$ replaced by ones in positions $(2,l_i^*)$, $i=1,\dots,p$. We show that all $l_i^* \le n$ and that E^* is a Pólya matrix. This can be done in steps, and we consider only the first one, which moves $e_{1,k_1}=1$ to $e_{2,l_1^*}=1$. If the resulting matrix is E_1, we compare the M_k-functions (see § 9) of E and E_1. If $k_1 = l_1^*$, the situation is trivial. Otherwise we have $e_{2,k}=1$ for $k \in I := [k_1, l_1^*)$, hence $M_k(E) \ge k+2$ in this range. Hence $l_1^* \le n$. Also, $M_k(E_1) = M_k(E)$, $k \notin I$, and $M_k(E_1) = M_k(E)-1$, $k \in I$. Thus, also E_1 satisfies the Pólya condition.

The set l_i^*, $i=1,\dots,p$, is minimal in the following sense. If the integers l_i satisfy $k_i \le l_i$ and $e_{2,l_i}=0$ for all $i=1,\dots,p$, and if $0 \le l_1 < \dots < l_p$, then $l_i \ge l_i^*$, $i=1,\dots,p$. This easily follows by induction on i. In this situation, if the shift of k_1 to $l_1,\dots$, of k_p to l_p has order α, then $\sum(l_i-k_i) = \sum(l_i^*-k_i) = \alpha$ and consequently $l_i = l_i^*$, $i=1,\dots,p$.

Obviously, there exists a multiple shift Λ_0^α of order α so that $\Lambda_0^\alpha E$ has no collision in rows 1 and 2. It moves $e_{1,k_p}=1$ to position $(1,l_p^*)$, then $e_{1,k_{p-1}}=1$ to $(1,l_{p-1}^*)$, and so on. From the minimality of $l_1^*\dots l_p^*$, any other shift Λ^α of order α which has no collisions, produces the same matrix $\Lambda^\alpha E = \Lambda_0^\alpha E$.

If $X^* = (x_2,\dots,x_m)$ and if the second row of E has ones in the positions $k_1^*,\dots k_q^*$, then

$$D(\Lambda_0^\alpha E, X)_{x_1=x_2} = (-1)^\sigma D(E^*,X^*)\,, \tag{10.3}$$

where σ is the number of permutations needed to bring the sequence $l_1^*,\dots,l_p^*$, $k_1^*,\dots,k_q^*$ to its natural order.

All terms of the sum (10.2) without collision are equal to (10.3). However, the shift Λ_0^α has in general many representations as a product of simple shifts. If their number is C, then

$$\frac{d^\alpha}{dx_1^\alpha}D\Big|_{x_1=x_2} = \varepsilon C D(E^*,X^*)\,. \tag{10.4}$$

By the inductive assumption, the last determinant is not identically zero, hence also $D(E,X)$ has this property. □

No equally simple necessary and sufficient conditions are known for the regularity of E. However, the following simple sufficient condition is often useful.

A sequence $e_{i,k} = \dots = e_{i,k+r-1} = 1$ in E is an *odd sequence* if r is odd and if $e_{i,k-1}=0$ (or $k=0$) and $e_{i,k+r}=0$ (or $k+r-1=n$). The *sequence is supported* if there exist ones $e_{i_1,k_1} = e_{i_2,k_2} = 1$ with $k_1,k_2 < k$ and $i_1 < i < i_2$. Thus, no sequence in rows $i=1$ or $i=m$ is supported, nor is any of the sequences which begin in column 0 or end in column n.

The proofs that follow will depend on a variation of Rolle's theorem, associated with an interpolation matrix E. For simplicity, this will be formulated for functions f that are analytic on a subinterval of $\mathbb{R}$. First, we note:

(10.5) *Between any two adjacent zeros $\alpha < \beta$ of f there is (counting multiplicities) an odd number of zeros of f'.*

Indeed, on (α,β) the function f is of constant sign; let, for example, $f(x)>0$. Then necessarily $f'(x)>0$ for x close to α, $x>\alpha$; and $f'(x)<0$ for x close to β, $x<\beta$. Thus, f' changes sign on (α,β). At its zeros of odd or of even order, f' does or does not change sign, respectively. Hence f' has an odd number of zeros of odd order on (α,β).

We shall say that the pair E,X annihilates f if

$$f^{(k)}(x_i) = 0 \quad \text{for} \quad e_{i,k}=1\,. \tag{10.6}$$

Theorem 10.2. *Let f be an analytic function annihilated by E,X where E is a Pólya matrix without odd supported sequences. Then there exist ξ_k, $k=0,\dots,n$, with $x_0 \le \xi_k \le x_m$ for which*

$$f^{(k)}(\xi_k) = 0\,, \quad k=0,\dots,n\,. \tag{10.7}$$

Proof. The theorem is obviously true if $n = 0$. In the general case, we prove the theorem by induction on n.

Because of the Pólya condition, there is a zero ξ_0 of f. It will be sufficient to find an interpolation matrix E' with columns $k = 1, \ldots, n$ which is a Pólya matrix without odd supported sequences and the corresponding set X' so that E', X' annihilate f'. We construct E', X' as follows:

(a) If E has a single one in column 0, E' is obtained from E by deleting this column; X' is the subset of the x_i involved in (10.6) with $k \geq 1$.

(b) If E has $t > 1$ ones in column 0, we consider an arbitrary pair $x_i < x_j$ of consecutive zeros of f from (10.6). Since f is analytic, we can choose zeros α, β with $x_i \leq \alpha < \beta \leq x_j$ such that f has no zeros on (α, β). By (10.5), f' has an odd number of zeros in (α, β) counting multiplicities. Hence there is a zero $\xi \in (\alpha, \beta)$ with odd multiplicity. If ξ does not appear as a zero of f' in (10.6), then $f'(\xi) = 0$ is a new interpolation condition for f'. If on the other hand $f'(\xi) = \cdots = f^{(r)}(\xi) = 0$ appear in (10.6) but not $f^{(r+1)}(\xi) = 0$, then r is even since E has no odd supported sequences. Hence $f^{(r+1)}(\xi) = 0$ is a new interpolation condition not listed by (10.6). We can assume that $r+1 \leq n$, for otherwise ξ will give all additional zeros required for (10.7). We then let E' be the new matrix formed from E by deleting column 0 and inserting the $t - 1$ new interpolation conditions. The set X' is obtained from X by omitting the x_i which appear in (10.6) only for $k = 0$, and adding the new ξ's. Induction completes the proof. □

The following theorem has been established by Atkinson and Sharma [1969] and Ferguson [1969]. It follows also from the theorem of Birkhoff about interpolation kernels (see Notes).

Theorem 10.3. *A Pólya interpolation matrix is regular if it has no odd supported sequences.*

Proof. It is sufficient to show that a polynomial P_n of degree $\leq n$ which satisfies the homogeneous conditions $P_n^{(k)}(x_i) = 0$, $e_{i,k} = 1$, is identically zero. But this follows from (10.7). □

Example. (a) An Hermite interpolation matrix has only rows of type $(1, \ldots, 1, 0, \ldots, 0)$. It has no supported sequences and is regular. This is also true if in a Pólya matrix only the *interior rows* $1 < i < m$ are assumed to be of this type.

(b) A two row matrix is regular if and only if it satisfies the Pólya conditions.

§ 11. Problems

11.1. For each trigonometric polynomial T of degree $< n$ one has

$$\frac{1}{2\pi}\int_0^{2\pi} T(x)\,dx = \frac{1}{n}\sum_1^n T\left(\frac{2\pi k}{n}\right).$$

11.2. If $|P_n(z)| \leq M$ in $|z| \leq 1$, then $|P_n(z)| \leq Mr^n$ in $|z| \leq r$ for $r > 1$.

11.3. If P_n is a polynomial with positive coefficients in x, $1-x$, then also $P_{n+1}(x) = \int_0^x P_n(t)\,dt$ has the property.

11.4. If $P(x) = \prod_1^n (x - z_k)$, where z_k can be complex, then on $[-1, 1]$, $\|P_n\|_\infty \geq 2^{-n+1}\prod_{|z_k|>1} |z_k|$ (Timan).

11.5. If for a real polynomial P_n, $\|P_n\| \leq M$ on $[-1, 1]$, then for each k, $|P_n^{(k)}| \leq M|C_n^{(k)}(x)|$, $x \notin [-1, 1]$, where C_n is the Chebyshev polynomial. For $k = 0$, this is Proposition 2.3.

11.6. If $f(x) = 1/x$ and $0 < x_0 \leq \cdots \leq x_n$, then $[x_0, \ldots, x_n]f = \frac{(-1)^n}{x_0 \cdots x_n}$.

11.7. If the points $x_0, \ldots, x_n$ are inside of the simple closed curve C and f is analytic on and inside C, then

$$[x_0, \ldots, x_n]f = \frac{1}{2\pi i}\int_C \frac{f(z)\,dz}{(z - x_0)\cdots(z - x_n)}.$$

11.8. From the formula given above derive the coefficients $\alpha_{i,l}$ of the relation $[x_0, \ldots, x_n]f = \sum \alpha_{i,l} f^{(l)}(x_i)$, mentioned in § 7.

11.9. Show that the proof of the continuity theorem of § 6 can be carried out also for Birkhoff interpolation, if E is a Pólya matrix with no odd supported sequences.

11.10. Show that the determinant $D(E, X)$ of (9.2) is a homogeneous polynomial in $x_1, \ldots, x_m$ of total degree $1 + 2 + \cdots + n - \sum_{e_{i,k}=1} k$.

11.11. Let β be the maximal order of shifts of row 1 of E which preserve the Pólya property of E. Then $D(E, X)$ is of degree $\leq \beta$ in x_1.

§ 12. Notes

12.1. There are many inequalities of Bernstein and Markov types. We can mention only some of them.

(a) Stechkin [1948] and Nikolskii have noticed that $T_n^{(k)}(x)$ in Bernstein's inequality can be replaced by $\Delta_h^k(T_n, x)/h^k$. They prove

$$\|T_n^{(k)}\|_\infty \leq \left(\frac{n}{2\sin nh}\right)^k \max_x |\Delta_h^k(T_n, x)| \leq n^k \|T_n\|_\infty. \tag{12.1}$$

(b) The best possible inequality for $\|P_n^{(k)}\|$ on $[-1, 1]$ has been found by W.A. Markov in 1892:

$$\|P_n^{(k)}\|_\infty \leq C_n^{(k)}(1)\|P_n\|_\infty, \tag{12.2}$$

where C_n stands for the n-th Chebyshev polynomial and

$$C_n^{(k)}(1) = \frac{n^2(n-1)^2\cdots(n^2-(k-1)^2)}{1.3.5\cdots(2k-1)}.$$

(c) The bound for $\|P_n'\|_\infty$ in terms of $\|P_n\|_\infty$ depends very much on the distribution of zeros of the polynomial P_n. In addition to results of § 4, we have: If all zeros of P_n are in $|z| \le 1$, then $\|P_n'\| \ge \frac{n}{2}\|P_n\|$ (Turán [1939]); if they are all outside of this disk, then $\|P_n'\| \le \frac{n}{2}\|P_n\|$ (Lax [1944]).

(d) Among the newer results we note that Bojanov [1982] has proved, for $P \in \mathcal{P}_n$ and the Chebyshev polynomial C_n, a generalization of (1.2):

$$\|P_n'\|_{L_p[-1,1]} \le \|C_n'\|_{L_p[-1,1]}\|P_n\|_{C[-1,1]}, \quad 1 \le p \le \infty.$$

For an encyclopedic review of inequalities of the types of Bernstein and Markov see the book of Rahman and Schmeisser [A-1983].

12.2. The divergence properties of Lagrange interpolation polynomials $P_n(f)$ are very pronounced for the interpolation points that are equally distributed in $[-1,1]$. The polynomials $P_n(f)$ can diverge on parts of the interval even if $f(x)$ is analytic on $[-1,1]$ (for example, if $f(x) = (x^2+a^2)^{-1}$ with small a (Runge)); moreover, (see Natanson [A-1965, vol. 3, p. 30]) Bernstein showed that they diverge everywhere except for $x = 0$ if $f(x) = |x|$.

12.3. One can achieve uniform convergence $P_n(f) \to f$ for all continuous functions f if one uses only a part $\le qn$, $0 < q < 1$ of the points $x_k^{(n)}$ for interpolation, and the rest for smoothing of the polynomial $P_n(f)$. In the Hermite-Fejér interpolation at the zeros $x_k^{(n)}$, $k = 1, \dots, n$, of C_n, one defines the polynomial $H_n(f,x) := H_n(x)$ by the conditions $H_n(x_k^{(n)}) = f(x_k^{(n)})$, $H_n'(x_k^{(n)}) = 0$, $k = 1, \dots, n$. One has the simple formula

$$H_n(f,x) = \sum_{k=1}^{n} f(x_k^{(n)})(1 - x x_k^{(n)})\left(\frac{C_n(x)}{n(x - x_k^{(n)})}\right)^2 \tag{12.6}$$

and the uniform convergence $H_n(f,x) \to f(x)$, $-1 \le x \le 1$.

12.4. The following inequality of Turán (for the proof see Landau-Gaier [B-1986, p. 170]) uses in its estimate *the number of terms* of a polynomial, rather than its *degree*. Let A_δ be an arc of the circle $\mathbb{T} := \{z : |z| = 1\}$ of length $\delta > 0$. If $\|\cdot\|_\delta$ and $\|\cdot\|$ stand for the uniform norm on A_δ and on $\mathbb{T}$, respectively, then for some absolute constant C,

$$\|P\|_\delta \le (C/\delta)^N \|P\|,$$

where N is the number of terms of P. This inequality has applications for estimates of analytic and of quasi-analytic functions.

12.5. Recently, T. Erdélyi has published (alone and in collaboration) several important polynomial inequalities, for example Erdélyi [1989] and Erdélyi, Máté and Nevai [1991].

12.6. Each polynomial $P_n(x)$ of degree n that satisfies $P_n(x) > 0$, $-1 < x < 1$, has a representation

$$P_n(x) = \sum_{j=0}^{N} a_j(1-x)^j(1+x)^{N-j} \quad \text{with all} \quad a_j \ge 0.$$

The smallest value of N here can be called the *Bernstein degree* $N(P)$ of P. Obviously, $N(P) \ge n$, but $N(P)$ can be much larger than n. According to Erdélyi and Szabados [1988] and Erdélyi [1991], if all zeros of $P \in \mathcal{P}_n$ are outside of the ellipse $x^2 + \varepsilon^{-2}y^2 = 1$, $0 < \varepsilon < 1$, then one has the *unimprovable inequality* $N(P) \ge Cn\varepsilon^{-2}$ for some absolute constant $C > 0$.

12.7. The history of Birkhoff interpolation is quite interesting. The paper [1906] of Birkhoff was written in 1905, when he was only 19 years old. His main interest was to find the representation of the remainder (12.7) and to study the properties of its kernel K. Thus he was the first to introduce what later became known as Peano kernels. Birkhoff gave a proof, albeit incorrect, of our Theorem 10.1. The Theorem 10.3 is not contained in [1906], but follows in an easy way from the properties of the kernel K, which he develops. The renewal of interest in Birkhoff interpolation started in 1966 with the paper of Schoenberg [1966]. He recognized the importance of the interpolation matrix E and of the Pólya condition.

12.8. For the interpolating polynomial P_n of (9.1) and the data $c_{i,k} = f^{(k)}(x_i)$, Birkhoff [1906] gives a formula for the remainder

$$f(x) - P_n(x) = \frac{1}{D(E,X)}\int_{x_1}^{x_m} f^{(n+1)}(t)K_E(X,t)\,dt. \tag{12.7}$$

The kernel $K_E(t) := K_E(X,t)$ – a piecewise polynomial (a spline) – has a simple representation by means of a determinant. The main theorem of Birkhoff states that the number of *changes of sign* of $K_E(t)$ for an indecomposable Pólya matrix E does not exceed the number of odd supported sequences of E. This has been amplified by Lorentz [1975]: the number of *zeros* of $K_E(t)$ does not exceed this bound.

12.9. In recent years, notions and methods of univariate Birkhoff interpolation have been used in the difficult problem of polynomial interpolation of functions of several variables. In contrast to single variables, there are here simple necessary and sufficient conditions for regularity (but very few regular methods), but nothing approaching a characterization of almost regularity. See for example R.A. Lorentz [A-1992].

Chapter 5. Splines

§ 1. Definitions and Simple Properties

Splines are piecewise polynomials with $\mathbb{R}$ as their natural domain of definition. We shall require that on each compact interval they consist of only finitely many polynomial pieces. Splines have been present in Analysis for a long time. They often appear as solutions to extremal problems. For example, the Euler splines introduced later in this chapter are extremal functions for both the Kolmogorov inequalities (in §6) and the Favard theorem (Chapter 7). Splines also appear as kernels of interpolation formulas.

The use of splines as approximants has received much impetus from the advent of computers. The systematic study of splines began with the work of I. J. Schoenberg in the forties. At present, splines are widely used in numerical computation and are indispensable for many theoretical questions. Schoenberg proved or conjectured many basic theorems of this theory.

Spline theory is the subject of many books, among them the introductory books of de Boor [A-1978] (which emphasizes the computational aspects) and [A-1990], of Nürenberger [A-1989] (emphasis on best approximation), Korneichuk [A-1984] and [A-1990] (problems where exact constants can be found). The book of Schumaker [A-1981] contains a wealth of material.

Let $T^* := (t_i^*)_1^s$ or $T^* := (t_i^*)_{-\infty}^{\infty}$ be a finite or biinfinite strictly increasing sequence of points of $\mathbb{R}$; in the second case, we assume that $|t_i^*| \to \infty$ for $i \to \pm\infty$. A function S on $\mathbb{R}$ is a *spline of order* r $(r = 1, 2, \dots)$, *equivalently of degree* $m = r - 1$, *with the breakpoints* T^* if on each interval (t_i^*, t_{i+1}^*), and on the intervals $(-\infty, t_1^*)$, (t_s^*, ∞) if T^* is finite, it is a polynomial of degree $\le m$, and on one of them of degree exactly m. At the breakpoints t_i^*, S and its derivatives (which are also splines) are defined by continuity, if possible, and not defined otherwise. Thus the splines of order one are step functions, those of order two are broken lines, and so on.

The *smoothness* m_i of a spline S at the breakpoint t_i^* is defined as follows: $m_i := 0$ if S is discontinuous at t_i^*. Otherwise, m_i is the largest integer $0 < m_i \le r$ so that S has continuous derivatives of orders $< m_i$, that is $S \in C^{(m_i - 1)}$ in some neighborhood of t_i^*. If $m_i = r$, then S is a polynomial in some neighborhood of t_i^*.

For given $A = [a, b]$ or $A = \mathbb{R}$ and T^* (in the first case we assume that $a < t_i^* < b$, $i = 1, \dots, s$) we form spline spaces on A. By $\mathcal{S}_r^*(A)$ and $\mathcal{S}_r^*(T^*, A)$, we denote the space of all splines of order $\le r$ on A, and of all such splines

whose breakpoints are contained in T^*, respectively. A set T^* and a sequence of integers m_i, $0 \le m_i < r$ define a *Schoenberg space* on A, which consists of all S of order $\le r$, with breakpoints contained in T^*, and of smoothness $\ge m_i$ at t_i^*. However, instead of the m_i one usually considers the *defect* $k_i := r - m_i$, since the latter describes the number of degrees of freedom of S at t_i^* (see Theorem 1.1). The Schoenberg space is then denoted $\mathcal{S}_r := \mathcal{S}_r(T^*, \boldsymbol{k}, A)$, $\boldsymbol{k} := (k_i)$. The space $\mathcal{S}_r(T^*, \mathbf{1}, A)$, with $k_i := 1$ for all i, is a subspace of any other Schoenberg space $\mathcal{S}_r(T^*, \boldsymbol{k}, A)$. The Schoenberg spaces are linear spaces; also if $S = \mathcal{S}_r(T^*, \boldsymbol{k}, A)$, then its integral S_1 belongs to the space $\mathcal{S}_{r+1}(T^*, \boldsymbol{k}, A)$ with the same $\boldsymbol{k}$. A Schoenberg space $\mathcal{S}_r(T^*, \boldsymbol{k}, A)$ contains the space P_{r-1} and at the other extreme is contained in $\mathcal{S}_r^*(T^*, A)$, which is nothing other than $\mathcal{S}_r(T^*, \boldsymbol{k}, A)$, with $k_i := r$ for all i.

A typical spline is the *truncated power* $(x - a)_+^k$, where

$$x_+^k := \begin{cases} x^k, & x \ge 0 \\ 0, & x < 0, \end{cases} \tag{1.1}$$

except that this is not defined for $x = 0$, $k = 0$. Each of the truncated powers $(x - t_i^*)_+^j$, $j = r - k_i, \dots, r - 1$ is in $\mathcal{S}_r(T^*, \boldsymbol{k}, A)$.

In the case $A = [a, b]$, by means of truncated powers, one can construct a natural basis for $\mathcal{S}_r(T^*, \boldsymbol{k}, A)$. A basis for a finite dimensional linear space Y consists of elements $S_1, \dots, S_n$ of Y such that each $S \in Y$ has the unique representation $\sum_{j=1}^N \alpha_j S_j$. The (uniquely defined) number N is the dimension of Y.

Often a basis can be found in the following way. We find elements $S_1, \dots, S_N$ and linear functionals $a_1, \dots, a_N$ on Y for which $a_j(S_i) = 0$, $i \ne j$, and $= 1$ if $i = j$ and so that $a_j(S) = 0$ for $j = 1, \dots, N$ implies $S = 0$. The a_j are called *dual functionals* for the S_j. One has in this case

$$S = \sum_1^N a_j(S) S_j\,. \tag{1.2}$$

Indeed, the result of applying a_i to each side is the same. Also (1.2) is the unique representation of S by the S_j since if $\sum_1^N b_j S_j$ is another representation, then applying a_i to both representations, we obtain $b_i = a_i(S)$.

Theorem 1.1. *If $A = [a, b]$ is a finite interval, then the space $\mathcal{S}_r(T^*, \boldsymbol{k}, I)$ has the basis*

$$\begin{aligned} S_{-j}(x) &:= (x - a)^j / j!\,, \quad j = 0, \dots, r - 1\,, \\ S_{i,j}(x) &:= (x - t_i^*)_+^j / j!\,, \quad j = r - k_i, \dots, r - 1\,, \quad i = 1, \dots, s\,, \end{aligned} \tag{1.3}$$

with the corresponding dual functionals

(1.4)

$$\begin{aligned} a_{-j}(S) &:= S^{(j)}(a)\,, & j &= 0, \dots, r - 1 \\ a_{i,j}(S) &:= S^{(j)}(t_i^* +) - S^{(j)}(t_i^* -)\,, & j &= r - k_i, \dots, r - 1\,, \quad i = 1, \dots, s\,. \end{aligned}$$

In particular,

$$\dim \mathcal{S}_r(T^*, \boldsymbol{k}, A) = n + r\,, \quad n := \sum_{i=1}^{r} k_i\,. \tag{1.5}$$

Proof. Let $S \in \mathcal{S}_r(T^*, \boldsymbol{k}, I)$. If $a_{i,j}(S) = 0$, $j = r - k_i, \dots, r-1$, $i = 1, \dots, p$, then $S, \dots, S^{(r-1)}$ are continuous at each t_i^*, hence $S^{(r-1)}$ is constant. Therefore, $S^{(r)} \equiv 0$ and S must be a polynomial of degree $< r$ on A. If in addition $a_{-j}(S) = 0$, $j = 0, \dots, r-1$, then S must be zero. The rest is obvious. □

For $S \in \mathcal{S}_r$, we have proved the representation

$$S(x) = P_{r-1}(x) + \sum_{i=1}^{s} \sum_{j=1}^{k_i} c_{i,j}\,(x - t_i^*)_+^{r-j} \tag{1.6}$$

which is sometimes taken as the definition of the space $\mathcal{S}_r$. In §3, we will introduce another, less obvious basis of B-splines, which is indispensable for the proof of approximation theorems by splines. Some properties of splines are best derived from (1.6), or directly from the definition of the space $\mathcal{S}_r$. Some examples are given in §7.

Splines (1.6) for which $S^{(m)}(x) \equiv 1$, $m = r-1$, are called *monosplines.* This happens when $P_{r-1} = \frac{x^m}{m!} + P_{r-2}$, and all $c_{i,1} = 0$, that is, when $S(x) = \frac{x^m}{m!} + S_1(x)$, $S_1 \in \mathcal{S}_{r-1}(T^*, \boldsymbol{k}, A)$. Splines with $|S^{(m)}(x)| = 1$ a.e., are called *perfect splines.* See §5 for examples.

The Schoenberg space $\mathcal{S}_r := \mathcal{S}_r(T^*, \boldsymbol{k}, I)$ contains functions which vanish on intervals, therefore it is not a Haar space. It is however a weak Haar space (§12, Chapter 3) of dimension $n + r$. This means that no spline $S \in \mathcal{S}_r$ can have $n+r$ changes of sign: there cannot exist points $x_1 < \cdots < x_{n+r+1}$, where S is defined, for which $S(x_k)S(x_{k+1}) < 0$, $k = 1, \dots, n+r$. See §§8, 10.

Splines satisfy elementary inequalities similar to polynomials. We mention only the following variants of the Nikolskii and Markov inequalities which hold for splines in the Schoenberg space $\mathcal{S}_r(T^*, \boldsymbol{k}, A)$, $A = [a, b]$, with breakpoints $T^* := \{t_j^*\}_1^s$. We define $t_0^* := a$, $t_{s+1}^* := b$.

Theorem 1.2. *If the breakpoints T^* satisfy*

$$\delta_0 \le |t_{j+1}^* - t_j^*| \le \delta\,, \quad j = 0, \dots, s\,, \tag{1.7}$$

one has for $S \in \mathcal{S}_r(T^*, \boldsymbol{k}, A)$, $A = [a, b]$,

$$\|S\|_p \le C\delta^{1/q - 1/p}\|S\|_q\,, \quad p_0 \le q \le p \le \infty \tag{1.8}$$

with $C = C(p_0, r)$; and for $k = 1, \dots, r-1$,

$$\|S^{(k)}\|_p \le C\delta_0^{-k}\|S\|_p\,, \quad 0 < p \le \infty \tag{1.9}$$

with $C = C(r)$.

Proof. This follows from (2.14) and (1.2) of Chapter 4. For instance, we shall prove (1.8). If $I_j := [t_j, t_{j+1}]$, $j = 0, \dots, s$, then since S is a polynomial of degree $< r$ on each I_j, we have from (2.14) of Chapter 4,

$$\begin{aligned} \|S\|_p^p &= \sum_{j=0}^{s} \|S\|_p(I_j)^p \le C \sum_{j=0}^{s} |I_j|^{p(1/q-1/p)} \|S\|_q(I_j)^p \\ &\le C\delta^{p(1/q-1/p)} \sum_{j=0}^{s} \|S\|_q(I_j)^p \end{aligned}$$

and since an l_p norm does not exceed a l_q norm, this last expression does not exceed $C\delta^{p(1/q-1/p)}(\sum \|S\|_q(I_j)^p)^{p/q}$ which gives (1.8). □

§ 2. B-Splines

The representation in Theorem 1.1 of a spline as a sum of a truncated powers is generally not very useful because the truncated power functions have large support. Curry and Schoenberg [1966] have stressed the importance of another basis which is more local in character. Their basis uses B-splines (basic splines) which are splines with the smallest possible support. They are defined by means of the divided differences of Chapter 4.

If $X : x_0 \le x_1 \le \cdots \le x_r$ is a sequence of $r+1$ points with $x_r \ne x_0$, we define the B-spline M by means of

$$M(x) := M(x; x_0, \dots, x_r) := r[x_0, \dots, x_r](\cdot - x)_+^{r-1}\,. \tag{2.1}$$

The points in X are called the *knots* of M. From §7 of Chapter 4, it follows that M is a linear combination of the truncated powers $(x_i - x)_+^{r-l_i}$ where l_i is the number of $x_j = x_i$, $j \le i$. Therefore, M is a spline function. Furthermore, since any r-th order divided difference of a polynomial of degree $r-1$ is zero, M vanishes identically when $x < x_0$ or $x > x_r$. From (7.16) of Chapter 4, we know that $M/r!$ is the Peano kernel for the divided difference at $x_0, \dots, x_r$: for any $f \in W_1^r$,

$$[x_0, \dots, x_r]\,f = \frac{1}{r!}\int_{-\infty}^{\infty} f^{(r)}(t)\,M(t)\,dt\,. \tag{2.2}$$

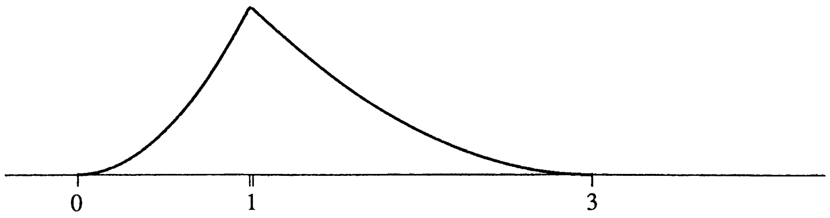
Putting here $f(t) = t^r$, we obtain

$$\int_{-\infty}^{\infty} M(t)\,dt = 1\,. \tag{2.3}$$

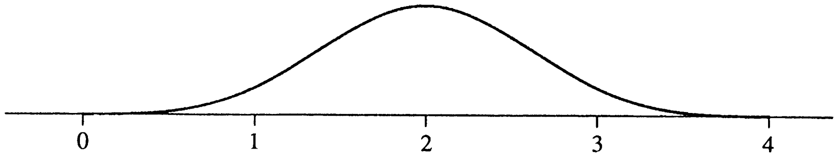
When $r = 1$, the B-spline is simply a multiple of the characteristic function χ of (x_0, x_1):

$$M(x; x_0, x_1) = \frac{1}{x_1 - x_0}\chi_{(x_0, x_1)}(x)\,, \quad x \ne x_0, x_1\,. \tag{2.4}$$

Other examples are the "hat function" for $r = 2$ and the continuously differentiable piecewise quadradic for $r = 3$ (see Figure 2.1).



The quadratic B-spline $M(0,1,1,3;\cdot)$ with knots $0,1,1,3$.



The cubic B-spline $M(0,1,2,3,4;\cdot)$ with knots $0,1,2,3,4$.

Fig. 2.1. Graphs of B-splines

Important properties of B-splines can be derived from properties of divided differences. As an example, we derive the following recurrence formula (Cox [1971] and de Boor [1972]) which is valid when $r \geq 2$,

$$\begin{aligned} &M(x;x_0,\dots,x_r) \\ (2.5)\qquad &= \frac{r}{r-1}\left[\frac{x-x_0}{x_r-x_0}M(x;x_0,\dots,x_{r-1}) + \frac{x_r-x}{x_r-x_0}M(x;x_1,\dots,x_r)\right] . \end{aligned}$$

Indeed, we apply to $(\cdot - x)(\cdot - x)_+^{r-2}$, the Leibnitz formula (7.10) of Chapter 4 and obtain

$$M(x;x_0,\dots,x_r) = r(x_0-x)[x_0,\dots,x_r](\cdot - x)_+^{r-2} + r[x_1,\dots,x_r](\cdot - x)_+^{r-2} .$$

The first term on the right (see (7.5) of Chapter 4) is equal to

$$\frac{r}{r-1}\frac{x_0-x}{x_r-x_0}\left[M(x;x_1,\dots,x_r) - M(x;x_0,\dots,x_{r-1})\right] ,$$

and the second is $\frac{r}{r-1}M(x;x_1,\dots,x_r)$. This yields (2.5).

From (2.4) and (2.5), we obtain by induction:

$$(2.6)\qquad \begin{aligned} &\text{(i)} && M(x) > 0\,, \quad x \in (x_0,x_r)\,; \quad M(x) = 0\,, \quad x \notin [x_0,x_r] \\ &\text{(ii)} && \begin{cases} M(x) \sim (x-x_0)^{r-k_0} & \text{for } x \to x_0+\,; \\ M(x) \sim (x-x_r)^{r-k_r} & \text{for } x \to x_r- \end{cases} \\ &\text{(iii)} && M(x) \leq C_r/(x_r-x_0)\,, \qquad x \in \mathbb{R}\,, \end{aligned}$$

where k_0 and k_r are the multiplicities of x_0 and x_r respectively in X, and C_r is a constant depending only on r. Property (i) is easy if x is not a knot; it

holds also if x is a knot of multiplicity less than r. Indeed, let $\delta > 0$. In the induction hypothesis for (i), one can add the assumption that $M(x) \geq C$ if $x \in [x_0+\delta, x_r-\delta]$ is not a knot, for $\delta > 0$, with $C > 0$ depending on δ, $x_0,\dots,x_r$. By continuity, this will hold for any knot of multiplicity $< r$ in $[x_0+\delta, x_r-\delta]$. We leave the proofs of (ii) and (iii) to the reader.

The B-spline M was normalized to have integral one. Another important normalization is

$$(2.7)\qquad \begin{aligned} N(x;x_0,\dots,x_r) &:= \tfrac{1}{r}(x_r-x_0)\,M(x;x_0,\dots,x_r) \\ &= (x_r-x_0)\,[x_0,\dots,x_r]\,(\cdot - x)_+^{r-1} . \end{aligned}$$

The recurrence formula for the B-splines N has a simple form; from (2.5) and (2.7),

$$(2.8)\qquad \begin{aligned} N(x;x_0,\dots,x_r) &= \frac{x-x_0}{x_{r-1}-x_0}N(x;x_0,\dots,x_{r-1}) \\ &+ \frac{x_r-x}{x_r-x_1}N(x;x_1,\dots,x_r) . \end{aligned}$$

It is also easy to differentiate the B-spline N. The operations of taking a divided difference with respect to t and taking a derivative with respect to x commute for the function $(t-x)_+^{r-1}$. Since $(x_r-x_0)\,[x_0,\dots,x_r]\,f = [x_1,\dots,x_r]\,f - [x_0,\dots,x_{r-1}]\,f$, we have for any x that is not a knot

$$(2.9)\qquad \begin{aligned} N'(x;x_0,\dots,x_r) &= -(r-1)\left([x_1,\dots,x_r] - [x_0,\dots,x_{r-1}]\right)(\cdot - x)_+^{r-2} \\ &= (r-1)\left(\frac{N(x;x_0,\dots,x_{r-1})}{x_{r-1}-x_0} - \frac{N(x;x_1,\dots,x_r)}{x_r-x_1}\right) . \end{aligned}$$

This formula also holds when $x = x_j$ is a knot provided x_j appears at most $r-2$ times in $x_0,\dots,x_r$ because $(t-x)_+^{r-2}$ is $r-3$ times continuously differentiable as a function of t.

Especially important are the B-splines with integer knots. Then $M_1(x) = \chi_I$, $I=[0,1]$ and $N_r := M_r := M(x;0,\dots,r)$ can be obtained by convolution:

$$(2.10)\qquad M_r = M_{r-1} * \chi_I = \chi_I * \cdots * \chi_I \quad r \text{ times} .$$

Indeed, $N_{r-1} * \chi_I$ has derivative $N_{r-1}(x) - N_{r-1}(x-1)$ which is the right side of (2.9) (for $x_0 = 0,\dots,x_r = r$).

In some applications, it is important to know how the B-spline $M(x;x_0,x_1,\dots,x_r)$ varies as a function of the knots $x_0,\dots,x_r$.

Lemma 2.1. *Suppose $(y_0,\dots,y_r)$ converges (in $\mathbb{R}^{r+1}$) to $(x_0,\dots,x_r)$ and A is a compact subset of $\mathbb{R}$.*

(i) *If $x_0,\dots,x_r$, are points from $\mathbb{R}$ and if of those from A not more than $r-1$ coincide, then $M(x;y_0,\dots,y_r)$ converges to $M(x;x_0,\dots,x_r)$ uniformly in $x \in A$. Similarly for $N(x;y_0,\dots,y_r)$.*

(ii) *If $0 < q < \infty$, then $N(x:y_0,\dots,y_r)$ converges to $N(x:x_0,\dots,x_r)$ in the metric of $L_q(A)$.*

Proof. From property (f) of §7, Chapter 4, the function $F(u_0,\dots,u_r) := r[u_0,\dots,u_r]t_+^{r-1}$ is continuous on any compact subset B of $\mathbb{R}^{r+1}$ which does not contain a point with r coordinates equal to zero. Since $M(x;y_0,\dots,y_r) = F(y_0-x,\dots,y_r-x)$, for all sufficiently small $\delta>0$, the set $B := \{(y_0-x,\dots,y_r-x) : x\in A \text{ and } |y_i-x_i|\le\delta\}$ has this property in case (i). Therefore the uniform convergence on A in case (i) follows.

By (2.6) (iii), the functions $N(x:y_0,\dots,y_r) = (y_r-y_0)M(x;y_0,\dots,y_r)$ are bounded and hence from (i) we obtain (ii). □

§ 3. B-Spline Series

The bases for the Schoenberg spaces $\mathcal{S}_r(T^*,\boldsymbol{k},A)$ discussed in §1 (which consist of powers and truncated powers) are quite simple and lead to simple formulas for the coefficients. Nevertheless, they have some serious disadvantages. The computation of a spline S by means of formula (1.6) requires the knowledge of S at all the breakpoints t_j^*. For most problems of approximation or interpolation, much more useful is the basis of B-splines, which is local in character. It allows us to compute $S(x)$ from data relating to S at only $r+1$ points near x. However, we must be prepared to pay the price of more complicated formulas for the coefficients (compare Theorem 3.2).

We shall adjust the notation for Schoenberg spaces $\mathcal{S}_r := \mathcal{S}_r(T^*,\boldsymbol{k},A)$ to reflect the new situation. Let $T^* = (t_i^*)$, $\boldsymbol{k} = (k_i)$. Each breakpoint t_i^* of multiplicity k_i, we now repeat k_i times. This gives an increasing sequence $T=(t_i)$. We call T the sequence of *basic knots* of the Schoenberg space, which we now denote by $\mathcal{S}_r := \mathcal{S}_r(T,A)$. For $A=\mathbb{R}$, we have $|t_i|\to\infty$, if $|i|\to\infty$, and in all cases (since $k_i\le r$)

$$t_i < t_{i+r}, \quad \text{for all } i. \tag{3.1}$$

If $A=[a,b]$, we shall also need *auxiliary knots* $t_{-r+1}\le\cdots\le t_0\le a$ and $t_{n+r}\ge\cdots\ge t_{n+1}\ge b$. Knots of multiplicity $k_i=1$ are called simple.

According to §1, the space $\mathcal{S}_r(T,A)$ consists of all splines S of order $\le r$ which at each *basic knot* of multiplicity k_i have continuous derivatives of orders $< r-k_i$. The basic knots completely determine $\mathcal{S}_r(T,A)$. The auxiliary knots are only needed for the construction of B-spline bases.

To a given knot sequence T corresponds the sequence of B-splines:

$$N_j(x) := N_{j,r}(x) := N_{j,T}(x) := N(x;t_j,\dots,t_{j+r}) \quad j\in\Lambda, \tag{3.2}$$

where $\Lambda := \mathbb{Z}$ if $A=\mathbb{R}$, $\Lambda := \{-r+1,\dots,n\}$, if $A=[a,b]$. It follows from §7, Chapter 4 that if $A=\mathbb{R}$, then N_j is a linear combination of the truncated powers $(t_i-x)_+^k$, $k=r-k_i,\dots,r-1$, and therefore $N_j\in\mathcal{S}_r(T,A)$. When $A=[a,b]$, one needs also truncated powers corresponding to the auxiliary knots. But these are polynomials of degree $<r$ on A and so $N_j\in\mathcal{S}_r(T,A)$ in this case as well.

The important Curry-Schoenberg theorem that follows says that the B-splines N_j, $j\in\Lambda$, are a basis for $\mathcal{S}_r := \mathcal{S}_r(T,A)$. For $A=\mathbb{R}$, as well as for $A=[a,b]$, this means that each spline $S\in\mathcal{S}_r$ is uniquely representable as a B-*spline series:*

$$S(x) = \sum_{j\in\Lambda} c_j N_j(x)\,. \tag{3.3}$$

If $x\in[t_i,t_{i+1}]$, only the terms $j=i-r+1,\dots,i$ in (3.3) are non-zero. Hence, each series (3.3) converges and its sum S is in $\mathcal{S}_r$.

Theorem 3.1 (Curry-Schoenberg [1966]). *The B-splines $(N_i)_{i\in\Lambda}$ with $\Lambda = \{-r+1,\dots,n\}$ and with $\Lambda=\mathbb{Z}$ are bases for the spaces $\mathcal{S}_r(T,A)$, $A=[a,b]$, and $A=\mathbb{R}$, respectively.*

We shall give two proofs of this theorem. The one in this section follows from the powerful result of de Boor-Fix [1973] (Theorem 3.2). A second completely independent proof based on zero counts for splines is given in §8.

As a preparation for Theorem 3.2, we consider some examples of (3.3). Perhaps the most important of these is the identity

$$\sum_{j\in\Lambda} N_j(x) = 1, \quad x\in A\,. \tag{3.4}$$

This shows that the N_j are a partition of unity on A and explains their normalization. If $A=[a,b]$, (3.4) is obtained by writing $N_j(x) = ([t_{j+1},\dots,t_{j+r}] - [t_j,\dots,t_{j+r-1}])(\cdot - x)_+^{r-1}$ and taking into account the relations

$$[t_{-r+1},\dots,t_0]\,(\cdot-x)_+^{r-1} = 0\,, \quad [t_{n+1},\dots,t_{n+r}]\,(\cdot-x)_+^{r-1} = 1\,, \quad x\in A\,.$$

The case $A=\mathbb{R}$ also follows from this.

We can use the recurrence formula (2.8) to find the B-spline series of the monomials $(\xi-x)^s$, $s=0,\dots,r-1$, $\xi\in\mathbb{R}$. These series are most easily expressed with the aid of the polynomials $g_j := g_{j,r}$ where $g_{j,1}(x) := 1$ and

$$g_{j,r}(x) := \frac{1}{(r-1)!}(x-t_{j+1})\dots(x-t_{j+r-1})\,, \quad r=2,3,\dots\,. \tag{3.5}$$

For example, we have the Marsden [1970] identities: for any $\xi\in\mathbb{R}$, $r=1,2,\dots$,

$$\frac{(\xi-x)^s}{s!} = \sum_{j\in\Lambda} g_{j,r}^{(r-s-1)}(\xi) N_{j,r}(x)\,, \quad x\in A\,, \quad s=0,\dots,r-1\,. \tag{3.6}$$

For the proof, we can restrict ourselves to the case $A=\mathbb{R}$. We consider first the special case $s=r-1$ of (3.6):

$$\frac{(\xi-x)^{r-1}}{(r-1)!} = \sum_{j\in\Lambda} g_{j,r}(\xi)N_{j,r}(x) \tag{3.7}$$

which we prove by induction on r. For $r = 1$, it follows from (3.4). For general r we use the recurrence formula (2.8) to replace the j-th term of the sum (3.7) by

$$g_{j,r}(\xi)\left(\frac{x-t_j}{t_{j+r-1}-t_j}N_{j,r-1}(x)+\frac{t_{j+r}-x}{t_{j+r}-t_{j+1}}N_{j+1,r-1}(x)\right).$$

Hence, we can rewrite the series (3.7) as a B-spline series of order $r-1$ where the coefficient of $N_{j,r-1}$ is

$$\frac{x-t_j}{(t_{j+r-1}-t_j)}g_{j,r}(\xi)+\frac{t_{j+r-1}-x}{(t_{j+r-1}-t_j)}g_{j-1,r}(\xi)=\frac{\xi-x}{r-1}g_{j,r-1}(\xi).$$

This shows that (3.7) is valid for $r = 1, 2, \dots$. To obtain (3.6), we differentiate the identity (3.7) $r - s - 1$ times with respect to ξ.

With the help of (3.6), we can find the B-spline series of *any polynomial* P of degree $\le r-1$ from its Taylor expansion at ξ:

$$P=\sum_{j\in\Lambda}c_j(P)N_j\,,\quad c_j(P)=\sum_{\nu=0}^{r-1}(-1)^\nu g_j^{(r-\nu-1)}(\xi)P^{(\nu)}(\xi)\,. \tag{3.8}$$

Different points ξ give the same value of $c_j(P)$ in (3.8). Indeed, if τ is any other point, we have $P^{(\nu)}(\xi)=\sum_{k=\nu}^{r-1}P^{(k)}(\tau)\frac{(\xi-\tau)^{k-\nu}}{(k-\nu)!}$. Substituting this into the sum (3.8) and rearranging terms, we obtain that this sum is equal to

$$\sum_{k=0}^{r-1}(-1)^kP^{(k)}(\tau)\sum_{\nu=0}^{k}g_j^{(r-\nu-1)}(\xi)\frac{(\tau-\xi)^{k-\nu}}{(k-\nu)!}\,.$$

The interior sum is the Taylor expansion of $g_j^{(r-k-1)}$. This establishes our statement.

We can now find the B-spline series of an arbitrary spline $S\in\mathcal{S}_r(T,A)$. We follow de Boor and Fix [1973] and define the linear functionals

$$c_j(S):=\sum_{\nu=0}^{r-1}(-1)^\nu g_j^{(r-\nu-1)}(\xi_j)S^{(\nu)}(\xi_j)\,,\quad S\in S_r(T,A)\,, \tag{3.9}$$

where the ξ_j are arbitrary points from $(t_j,t_{j+r})\cap A$. An explanation of this formula is needed when ξ_j is one of the knots t_i, $t_j<t_i<t_{j+r}$. If k_i is the multiplicity of t_i, then the derivatives $S^{(\nu)}(t_i)$, $r-k_i\le\nu\le r-1$ need not exist. But then $r-\nu-1\le k_i-1$, and since t_i is a zero of order k_i of g_j, we have $g^{(r-\nu-1)}(t_i)=0$. We shall therefore interpret the terms with $g_j^{(r-\nu-1)}(\xi_j)=0$ in (3.9) to be zero.

It will be established in the following theorem that the value of $c_j(S)$ does not depend upon the choice of ξ_j.

Theorem 3.2 (de Boor-Fix). *If $A=[a,b]$ or $A=\mathbb{R}$, then the B-splines N_j, $j\in\Lambda$, are a basis for $\mathcal{S}_r=\mathcal{S}_r(T,A)$: each $S\in\mathcal{S}_r$ can be written uniquely as a B-spline series:*

$$S(x)=\sum_{j\in\Lambda}c_j(S)N_j(x)\,,\quad x\in A\,, \tag{3.10}$$

where c_j, $j\in\Lambda$, are the de Boor-Fix functionals (3.9).

Proof. Since the case $A=\mathbb{R}$ follows from the compact interval case, we can assume that $A=[a,b]$. By our previous remarks, the representation (3.10) is valid when S is a polynomial of degree $<r$. We now check its validity when $S_1(x):=(x-t_i)_+^{r-k}$ where t_i is a knot in T of multiplicity at least k. For the polynomial $P(x):=(x-t_i)^{r-k}$, the coefficient $c_j(P)=0$ if $t_j<t_i<t_{j+r}$. Indeed, we can take $\xi:=t_i$ in (3.8), then g_j has a zero of multiplicity at least k at ξ while P has a zero of multiplicity $r-k$. Hence each term of the sum (3.8) is zero. From this we derive that $c_j(S_1)=0$, if $t_j<t_i<t_{j+r}$, for any choice of ξ_j. Indeed, if we take $\xi_j>t_i$, the derivatives $S_1^{(\nu)}(\xi_j)$ in (3.9) will be the same as the derivatives $P^{(\nu)}(\xi_j)$, and $c_j(S_1)=c_j(P)=0$. If we take $\xi_j<t_i$, we will have $c_j(S_1)=c_j(0)=0$. By continuity, this will even be true for $\xi_j=t_i$ because as we have observed in the sum (3.9) the coefficient of $S_1^{(\nu)}$ is zero if this derivative does not exist. Now, if $x<t_i$, any term in the sum (3.10) for S_1 for which $N_j(x)\ne0$ has zero coefficient while if $x>t_i$ then any such term has coefficient $c_j(P)$. We see therefore that (3.10) holds if $x\ne t_i$. By continuity, it also holds for $x=t_i$ provided $k<r$, that is, S_1 is continuous.

Since polynomials and truncated powers span $\mathcal{S}_r$, the representation (3.10) is valid for each $S\in\mathcal{S}_r$. This representation is unique since the dimension of $\mathcal{S}_r$ is equal to the number of the functions N_j. □

Another formulation of Theorem 3.2 is that the functionals c_j given by (3.9) are *dual* to the basis N_j : $c_j(N_i)=1$ if $i=j$, $=0$ otherwise.

When $A=[a,b]$, the right side of (3.10) gives an extension of S from A to all of $\mathbb{R}$. For different auxiliary knots we obtain different B-splines and therefore different extensions. The simplest of these is when all of the auxiliary knots are taken as the end points of A. In this case, S is defined by (3.10) to be zero outside A.

The B-splines are splines of smallest support. If $S\not\equiv0$ belongs to $\mathcal{S}_r(T,A)$ and *vanishes outside of* a finite interval I then I *contains at least* $r+1$ *knots*. Indeed, if this interval were to contain fewer than $r+1$ knots, then for each $j\in\Lambda$, we can choose the points ξ_j in (3.9) to be outside of I. Then $c_j(S)=0$, and by (3.10), $S\equiv0$. In particular, this shows that the B-splines N_j, $j\in\Lambda$, have the smallest possible support for splines in $\mathcal{S}_r(T,A)$.

It is very simple to differentiate the B-spline series (3.3): from (2.9) we get

$$S'(x)=(r-1)\sum_{j\in\Lambda}\frac{c_j-c_{j-1}}{t_{j+r-1}-t_j}N_{j,r-1}(x)\,. \tag{3.11}$$

This formula holds if x is not a knot of multiplicity $\ge r-2$.

§ 4. Quasi-Interpolant Operators

For splines S that belong to the Schoenberg space $\mathcal{S}_r := \mathcal{S}_r(T, A)$ where $A = [a, b]$ or $\mathbb{R}$, $T := (t_j)_{j\in\Lambda}$, $\Lambda := \{-r+1, \dots, n+r\}$ or $\Lambda := \mathbb{Z}$, we have the B-spline series of order r

$$S = \sum_{j\in\Lambda} c_j(S) N_j \,, \tag{4.1}$$

where the coefficients c_j are given by (3.9). The properties of this series will allow us to obtain good approximations for functions $f \in L_p(A)$, $1 \le p \le \infty$, by splines in $\mathcal{S}_r$.

This will be achieved by means of an operator Q (called a quasi-interpolant) which is a projection from L_1 onto $\mathcal{S}_r$ and thereby from each L_p onto $\mathcal{S}_r$ for each $1 \le p \le \infty$. Any projection Q onto $\mathcal{S}_r$ is necessarily of the form

$$Q(f) = \sum_{j\in\Lambda} a_j(f) N_j \,, \tag{4.2}$$

where the a_j are suitable linear functionals on L_1. Comparing this with (4.1), we see that $a_j(S) = c_j(S)$ must hold for each $S \in \mathcal{S}_r$; that is, a_j must be an extension of the functional c_j to all of L_1. The remarkable point is that we can choose the extensions a_j in such a way that the norm of the projection Q onto L_p can be estimated from above by a constant which does not depend on p nor on the knot sequence T.

The importance of the latter fact lies in the Lebesgue lemma (see (4.6) of Chapter 2) according to which the degree of approximation of $f \in L_p$ by $Q(f)$ satisfies

$$\|f - Q(f)\|_p \le (1 + \|Q\|) E(f, \mathcal{S}_r(T, A))_p \,. \tag{4.3}$$

The fact that there is an upper bound for $\|Q\|$ independent of p and T is much more than what is true for trigonometric or polynomial approximation, see Chapters 7, 8 and 10.

In order to define the quasi-interpolants, we shall need some properties of the B-spline series in the space $L_p(A)$. The first of these is a relation between the $L_p(A)$ norm of a spline $S \in \mathcal{S}_r(T, A)$ and its sequence $\boldsymbol{c} := (c_j)$ of B-spline coefficients. For example, if $A = \mathbb{R}$ and if $T = \mathbb{Z}$ are the integers, we shall prove that $\|S\|_p \sim \|\boldsymbol{c}\|_p$, that is

$$C_1 \|\boldsymbol{c}\|_p \le \|S\|_p \le C_2 \|\boldsymbol{c}\|_p \tag{4.4}$$

with C_1, C_2 depending only on r. Similar inequalities hold for general knot sequences T (Theorem 4.2 below).

In the construction of quasi-interpolants that follows, it will be convenient to require that the auxiliary knots be chosen with the following restrictions: $t_0 := a$, $t_{n+1} := b$ and

$$|t_{i+1} - t_i| \le |t_1 - t_0| \,, \text{ if } i < 0 \,, \quad |t_{i+1} - t_i| \le |t_{n+1} - t_n| \,, \text{ if } i > n \,. \tag{4.5}$$

We begin by estimating $c_j(S)$ of (3.9) by the norm of S. In what follows, and in Chapters 12 and 13, when dealing with B-spline series, *we shall always denote by J_j, the largest subinterval (t_i, t_{i+1}) of (t_j, t_{j+r}), contained in A*; from (4.5) we see that there are always such subintervals. If there are several such intervals, we take the one contained in A with the smallest i. It follows that $|J_j| \ge (t_{j+r} - t_j)/r$.

Lemma 4.1. *There is a constant $C > 0$ depending only on r (and p for p close to zero) with the property that for each $0 < p \le \infty$, and $S \in \mathcal{S}_r(T, A)$,*

$$|c_j(S)| \le C\, |J_j|^{-1/p} \|S\|_p (J_j) \,, \quad j \in \Lambda \,. \tag{4.6}$$

In addition, c_j depends only on the values of S in J_j.

Proof. We choose the point ξ_j of (3.9) to be the center of J_j. This will establish the last statement of the Lemma. From the Markov inequality (1.2) and inequality (2.14) of Chapter 4, for $0 < p \le \infty$,

$$\left|S^{(\nu)}(\xi_j)\right| \le C\, |J_j|^{-\nu} \|S\|_\infty (J_j) \le C |J_j|^{-\nu - 1/p} \|S\|_p (J_j) \,. \tag{4.7}$$

Moreover,

$$\left|g_j^{(r-\nu-1)}(\xi_j)\right| \le C \max_{j<i<j+r} |\xi_j - t_i|^\nu \le C |J_j|^\nu \,,$$

and (4.6) follows. ☐

We can now prove inequalities of the type (4.4). For general knot sequences, these inequalities must take into account the spacing of the knots T. For this purpose, we define the constants

$$d_j := \frac{t_{j+r} - t_j}{r} \,, \quad j \in \Lambda \,. \tag{4.8}$$

Theorem 4.2 (de Boor [1973_1]). *There is a constant $D_r > 0$ such that for each spline $S = \sum_{j\in\Lambda} c_j N_j$ of order r, and for each $0 < p \le \infty$,*

$$\begin{cases} D_r \|\boldsymbol{c}'\|_p \le \|S\|_p \le \|\boldsymbol{c}'\|_p \,, & 1 \le p \le \infty \\ D_r \|\boldsymbol{c}'\|_p \le \|S\|_p \le r^{1/p} \|\boldsymbol{c}'\|_p \,, & 0 < p < 1 \end{cases} \tag{4.9}$$

where $\boldsymbol{c}' := (c_j')$, $c_j' := c_j d_j^{1/p}$, $j \in \Lambda$.

Proof. Since $\sum_{j\in\Lambda} N_j \equiv 1$ on A, the right inequality in (4.9) is immediate for $p = \infty$. When $1 \le p < \infty$, we use Hölder's inequality to find

$$\left|\sum_{j\in\Lambda} c_j N_j\right| \le \left(\sum_{j\in\Lambda} |c_j|^p N_j\right)^{1/p} \left(\sum_{j\in\Lambda} N_j\right)^{1/p'} = \left(\sum_{j\in\Lambda} |c_j|^p N_j\right)^{1/p} .$$

Since $\int_{-\infty}^{\infty} N_j dt = d_j$, the right inequality in (4.9) follows. When $p < 1$, we have $\left|\sum_{j\in\Lambda} c_j N_j\right|^p \le \sum_{j\in\Lambda} |c_j|^p N_j^p$. Since $\int_{-\infty}^{\infty} N_j^p dt \le t_{j+r} - t_j = rd_j$, the right inequality follows in this case as well.

For the left inequalities, we derive from (4.6) that

$$|c_j'|^p \le C^p \int_{t_j}^{t_{j+r}} |S(x)|^p dx\,.$$

Since each $x \in A$ appears in at most r intervals (t_j, t_{j+r}), adding up these inequalities, we obtain the left estimates of (4.9). □

Corollary 4.3. *In the case $A = \mathbb{R}$, the B-splines series* (4.1) *converges to S for each $S \in \mathcal{S}_r(T,\mathbb{R}) \cap L_p$, $0 < p < \infty$, in the L_p metric.*

Proof. If $S \in \mathcal{S}_r(T,\mathbb{R}) \cap L_p$, then (4.9) yields that $\|\boldsymbol{c}'\|_p < \infty$. Then, as $m \to \infty$,

$$\left\| S - \sum_{-m}^{m} c_j N_j \right\|_p^p = \left\| \sum_{|j|>m} c_j N_j \right\|_p^p \le C \sum_{|j|>m} |c_j'|^p \to 0\,. \qquad \square$$

From Lemma 4.1, $p = 1$, we see that c_j is a bounded linear functional on $\mathcal{S}_r(T,A)$ in the L_1 norm. If γ_j is the Hahn-Banach extension of c_j onto $L_1(J_j)$, we will have

$$|\gamma_j(f)| \le C\,(t_{j+r} - t_j)^{-1}\,\|f\|_1\,(J_j)\,, \tag{4.10}$$

with C depending only on r. Then, Hölder's inequality yields for any $1 \le p \le \infty$,

$$|\gamma_j(f)| \le C\,(t_{j+r} - t_j)^{-1/p}\,\|f\|_p(J_j)\,. \tag{4.11}$$

We can now define *the quasi-interpolant of order r corresponding to the knots T in A*. For each locally integrable f it is given by

$$Q(f) := Q_T(f) := \sum_{j\in\Lambda} \gamma_j(f) N_j\,. \tag{4.12}$$

In particular, this is a projection from $L_p(A)$, $1 \le p \le \infty$, onto its subspace $\mathcal{S}_r(T,A)$. To estimate its norm, it is convenient to use, together with $[t_j, t_{j+1}]$, the slightly larger interval

$$I_j := [t_{j-r+1}, t_{j+r}] \cap A\,. \tag{4.13}$$

We also let Λ_j be the set of indices k for which N_k does not vanish identically on $[t_j, t_{j+1}]$:

$$\Lambda_j := \{j - r + 1, \dots, j\}\,. \tag{4.14}$$

Then, $J_k \subset I_j$, for all $k \in \Lambda_j$.

Theorem 4.4. *For each Schoenberg space $\mathcal{S}_r(T,A)$, $A = [a,b]$ or $\mathbb{R}$, for $f \in L_p(A)$, $1 \le p < \infty$, $f \in C(A)$, $p = \infty$, and for some constant C_r, one has the local and global norm estimates*

$$\begin{cases} \|Q_T(f)\|_p\,[t_j, t_{j+1}] \le C_r \|f\|_p(I_j)\,, \\ \|Q_T(f)\|_p(A) \le C_r \|f\|_p(A)\,. \end{cases} \tag{4.15}$$

Proof. For $x \in [t_j, t_{j+1}]$, we have $Q_T(f,x) = \sum_{k\in\Lambda_j} \gamma_k(f) N_k(x)$. Also, $\sum_{k\in\Lambda_j} N_k \equiv 1$ on this interval. Therefore, from (4.11) we obtain

$$\begin{aligned} \|Q_T(f)\|_p\,[t_j, t_{j+1}] &\le \max_{k\in\Lambda_j} |\gamma_k(f)| \left\| \sum_{k\in\Lambda_j} N_k \right\|_p [t_j, t_{j+1}] \\ &\le C \max_{k\in\Lambda_j} (t_{k+r} - t_k)^{-1/p}\,\|f\|_p(J_k)\,(t_{j+1} - t_j)^{1/p} \qquad (4.16) \\ &\le C\|f\|_p(I_j) \end{aligned}$$

since $J_k \subset I_j$ for the k in question. This gives the first inequality in (4.15) and also the second when $p = \infty$. When $1 \le p < \infty$, we raise both sides of the first inequality of (4.15) to the power p and sum over all $j \in \Lambda$. Since each $x \in A$ appears in at most $2r - 1$ of the intervals I_j, we obtain (4.15). □

Many interesting approximation properties of the Q_T follow from Theorem 4.4. For example from (4.3), we obtain that up to a constant factor C_r, the spline $Q_T(f)$ approximates f in the L_p norm as well as any other spline from $\mathcal{S}_r(T,A)$.

Another corollary is that one can estimate the global deviation $\|f - Q_T(f)\|(A)$ by means of the local errors $E_{r-1}(f, I_j)$ of approximating f by algebraic polynomial of degree $< r$.

Theorem 4.5. *For the quasi-interpolant* (4.12) *of order r on $A = [a,b]$ and each $f \in L_p(A)$,*

$$\|f - Q_T(f)\|_p\,(A) \le C_r \begin{cases} \left(\sum_{j=0}^{n} E_{r-1}(f, I_j)_p^p\right)^{1/p}, & 1 \le p < \infty \\ \max_{j=0,\dots,n} E_{r-1}(f, I_j)_p\,, & p = \infty\,. \end{cases} \tag{4.17}$$

Remark. The inequality (4.17) also follows for $0 < p < 1$ provided $Q_\Gamma(f)$ is defined and (4.15) holds for f and p.

Proof. The projection Q_T reproduces polynomials P of degree $< r$. We take for P the polynomial which realizes the error of approximation $E_r(f, I_j)_p$ and obtain for $x \in [t_j, t_{j+1}]$,

$$|f(x) - Q_T(f,x)| \le |f(x) - P(x)| + |Q_T(f - P, x)|\,.$$

Thus, from (4.15) of Theorem 4.4,

$$\|f - Q_T(f)\|_p\,[t_j, t_{j+1}] \le E_r\,(f, I_j)_p + C\|f - P\|_p(I_j)\,, \qquad (4.18)$$
$$\le C\,E_r\,(f, I_j)_p\,.$$

For $p = \infty$, this yields immediately (4.17). When $1 \le p < \infty$, we raise both sides of (4.18) to the power p and sum over all j. □

It is also simple to estimate the derivatives of $Q_T(f)$.

Proposition 4.6. *For the spline $S := Q_T(f)$, one has for $k = 1, \ldots, r-1$*

$$(4.19) \qquad \left|S^{(k)}(x)\right| \le C\,(t_{j+1} - t_j)^{-k-1/p}\,E_k(f, I_j)_p\,, \quad x \in (t_j, t_{j+1})\,.$$

Proof. If P is any polynomial of degree $< k$, then $S - P = Q_T(f - P) = \Sigma\gamma_j(f - P)N_j$ and therefore $S^{(k)} = \Sigma\gamma_j(f-P)N_j^{(k)}$. As in (4.16) we have for $x \in (t_j, t_{j+1})$,

$$\left|S^{(k)}(x)\right| \le \max_{\nu \in \Lambda_j} |\gamma_\nu(f - P)| \sum_{\nu \in \Lambda_j} \left|N_\nu^{(k)}(x)\right| \qquad (4.20)$$
$$\le C\,(t_{j+1} - t_j)^{-k-1/p}\,\|f - P\|_p(I_j)\,.$$

We have used here the Markov's inequality (1.2) of Chapter 4 on (t_j, t_{j+1}) to estimate $N_\nu^{(k)}(x)$. By taking an infimum over P, we obtain (4.19). □

§ 5. Euler and Bernoulli Splines

Splines often appear as solutions to extremal problems. Two important examples are the Euler splines which are extremal for the Kolmogorov-Landau inequalities of §7 and the Bernoulli splines which are used in the proof of the Favard theorem of Chapter 7. They share many properties: both are periodic and the spline of degree m can be obtained as a periodic integral of the spline of degree $m-1$.

The *Euler splines* $\mathcal{E}_m$ of degree $m = 0, 1, \ldots$ and period 2 are defined by

$$(5.1) \qquad \mathcal{E}_0(x) := (-1)^\nu\,, \quad \nu - \tfrac{1}{2} < x < \nu + \tfrac{1}{2}\,, \quad \nu \in \mathbb{Z}\,,$$
$$\mathcal{E}_1(x) := \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{E}_0(x + t)dt$$

and in general for $x \in \mathbb{R}$

$$(5.2) \qquad \mathcal{E}_{m+1}(x) := c_m \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{E}_m(x + t)dt\,, \quad c_m^{-1} = \int_{-\frac{1}{2}}^{\frac{1}{2}} \mathcal{E}_m(t)dt\,.$$

Thus $\mathcal{E}_1$ is a piecewise linear function which has breakpoints at the integers and values $\mathcal{E}_1(\nu) = (-1)^\nu$, $\nu \in \mathbb{Z}$. By induction, $\mathcal{E}_m$ is a spline of degree m and belongs to $C^{(m-1)}(\mathbb{R})$. It has simple knots at the integers when m is odd and at the "half integers" $v + \frac{1}{2}$, $v \in \mathbb{Z}$, when m is even. Other important properties of $\mathcal{E}_m$ are:

(5.3)
- (i) $\mathcal{E}_m$ is a periodic function of period 2, and $\mathcal{E}_m(x+1) = -\mathcal{E}_m(x)$.
- (ii) The integral of $\mathcal{E}_m$ over an interval of length 2 is zero.
- (iii) $\mathcal{E}_m(x)$ is an even and $\mathcal{E}_m\left(x + \frac{1}{2}\right)$ is an odd function of x.
- (iv) $\mathcal{E}_m(x)$ is positive for $x \in \left(-\frac{1}{2}, \frac{1}{2}\right)$.
- (v) $\mathcal{E}_m(x)$ is non-increasing for $x \in (0, 1)$.
- (vi) $\mathcal{E}_m(\nu) = (-1)^\nu\,, \quad \nu \in \mathbb{Z}$.
- (vii) $\|\mathcal{E}_m\|_\infty = 1$.

These properties can be proved by induction. If (5.3) holds for some integer m, then we show that it holds also for $m+1$. For example, we obtain $\mathcal{E}_{m+1}(x+1) = -\mathcal{E}_{m+1}(x)$ upon replacing $\mathcal{E}_m(x+1+t)$ by $-\mathcal{E}_m(x+t)$ in the integral (5.2). Since $\mathcal{E}'_{m+1}(x) = c_m\left(\mathcal{E}_m\left(x + \frac{1}{2}\right) - \mathcal{E}_m\left(x - \frac{1}{2}\right)\right) = 2c_m\mathcal{E}_m\left(x + \frac{1}{2}\right)$, by (iii), $\mathcal{E}'_{m+1}$ is odd and therefore $\mathcal{E}_{m+1}$ is even. Similar ideas prove the other statements in (5.3).

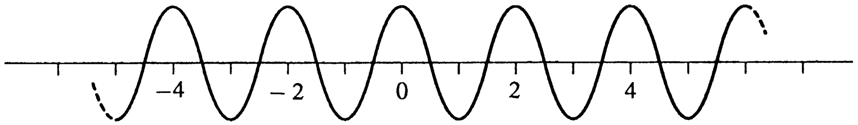


Fig. 5.1. The quadratic Euler spline $\mathcal{E}_2$

The periodic function $\mathcal{E}_m$ can be represented by its Fourier series:

$$(5.4) \qquad \mathcal{E}_m(x) = \frac{4}{\pi K_m} \sum_{j=0}^{\infty} \left[\frac{(-1)^j}{(2j+1)}\right]^{m+1} \cos(2j+1)\pi x\,, \quad m = 0, 1, \ldots\,,$$

where K_m are the so-called *Favard numbers*

$$(5.5) \qquad K_m := \begin{cases} \dfrac{4}{\pi} \displaystyle\sum_0^\infty \frac{1}{(2j+1)^{m+1}}\,, & \text{if } m \text{ is odd}\,, \\[2ex] \dfrac{4}{\pi} \displaystyle\sum_0^\infty \frac{(-1)^j}{(2j+1)^{m+1}}\,, & \text{if } m \text{ is even}\,. \end{cases}$$

For $m = 0$, this follows from the relation $K_0 = 1$, $\int_{-1}^1 \mathcal{E}_0(t)dt = 0$, and

$$\int_{-1}^1 \mathcal{E}_0(t)\cos k\pi t\,dt = 2\left[\int_0^{\frac{1}{2}} - \int_{\frac{1}{2}}^1\right] \cos k\pi t\,dt = \frac{4\sin k\pi/2}{k\pi}\,, \quad k = 1, 2, \ldots\,.$$

In this case, the series in (5.4) converges for all x which are not half integers. If $m \ge 1$, (5.4) is proved by induction. If we assume that this relation is true for $\mathcal{E}_{m-1}$, then (5.2) implies that (5.4) is valid up to some constant factor; since $\mathcal{E}_m(0) = 1$, this factor is $4/(\pi K_m)$.

The constants K_m can be rationally expressed in terms of π. We have $K_1 = \pi/2$, $K_2 = \pi^2/8$, $K_3 = \pi^3/24$, moreover $\lim_{m\to\infty} K_m = 4/\pi$ and

$$1 = K_0 < K_2 < \cdots < 4/\pi < \cdots < K_3 < K_1 = \pi/2\,. \tag{5.6}$$

Another useful identity concerns the derivatives of the Euler splines. For $0 \le k \le m$, we have

$$\mathcal{E}_m^{(k)}(x) = \frac{K_{m-k}}{K_m}\pi^k \begin{cases} (-1)^{k/2}\mathcal{E}_{m-k}(x)\,, & \text{if } k \text{ is even}\,, \\ (-1)^{(k-1)/2}\mathcal{E}_{m-k}(x+\frac{1}{2})\,, & \text{if } k \text{ is odd}\,, \end{cases} \tag{5.7}$$

provided x is not a breakpoint when $k = m$. This is proved by comparing terms in the Fourier series of $\mathcal{E}_m'$ and $\mathcal{E}_{m-1}$. In this way, we see that

$$\mathcal{E}_m'(x) = (\pi K_{m-1}/K_m)\,\mathcal{E}_{m-1}\left(x+\tfrac{1}{2}\right)\,. \tag{5.8}$$

Since $\mathcal{E}_m\left(x+\frac{1}{2}\right)$ is odd, we have $\mathcal{E}_m(x+1) = -\mathcal{E}_m(x)$. Thus (5.7) follows by a repeated application of (5.8).

The *Bernoulli spline* of degree 1 is defined by

$$\mathcal{B}_1(x) := (\pi - x)/2\,, \quad 0 \le x < 2\pi \tag{5.9}$$

and by 2π-periodicity on $\mathbb{R}$. For $m > 1$, $\mathcal{B}_m$ is defined inductively as the 2π-periodic integral of mean value 0 of $\mathcal{B}_{m-1}$. For example, for $x \in [0, 2\pi]$

$$\mathcal{B}_2(x) := -\frac{(\pi - x)^2}{4} + \frac{\pi^2}{12}\,. \tag{5.10}$$

If c_m is any zero of $\mathcal{B}_m$,

$$\mathcal{B}_m(x) = \int_{c_m}^{x} \mathcal{B}_{m-1}(t)dt\,, \quad m = 2, \ldots\,. \tag{5.11}$$

It follows that the $\mathcal{B}_m$ are splines of degree m which belong to $C^{(m-2)}$ and have breakpoints at the integral multiples of 2π. On $(0, 2\pi)$, $\mathcal{B}_m$ is a polynomial of degree m (the *Bernoulli polynomial*).

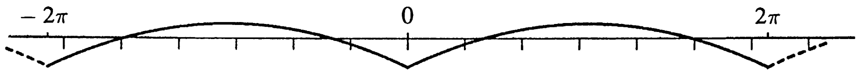


Fig. 5.2. The quadratic Bernoulli spline $\mathcal{B}_2$

For $\mathcal{B}_1$, we have the Fourier expansion

$$\mathcal{B}_1(x) = \sum_{j=1}^{\infty} \frac{\sin jx}{j}\,. \tag{5.12}$$

Integrating this, we obtain the representations

$$\mathcal{B}_m(x) = \sum_{j=1}^{\infty} j^{-m} \cos\left(jx - \frac{m\pi}{2}\right). \tag{5.13}$$

One of the motivations for the definition of $\mathcal{B}_1$ is that $d\mathcal{B}_1/\pi = d\rho - dt/2\pi$, where $d\rho$ is the Dirac measure with unit mass at 0. From this, we derive:

Theorem 5.1. *Each function $f \in W_p^m(\mathbb{T})$, $m = 1, 2, \ldots$, $1 \le p \le \infty$, with mean value 0 has a representation $f = \phi * \mathcal{B}_m$, that is,*

$$f(x) = \frac{1}{\pi}\int_{\mathbb{T}} \phi(t)\mathcal{B}_m(x-t)dt \tag{5.14}$$

with a unique function $\phi \in L_p(\mathbb{T})$, namely with $\phi = f^{(m)}$.

Proof. Since f has mean value 0, $\frac{1}{\pi}\int_{\mathbb{T}} f(t)d\mathcal{B}_1(x-t)dt = f(x)$. Therefore, from the periodicity of f and of the Bernoulli splines, we obtain by partial integration:

$$\begin{aligned} f(x) &= \frac{1}{\pi}\int_{\mathbb{T}} f(t)d\mathcal{B}_1(x-t)\,dt = \frac{1}{\pi}\int_{\mathbb{T}} f'(t)\mathcal{B}_1(x-t)\,dt \\ &= -\frac{1}{\pi}\int_{\mathbb{T}} f'(t)d\mathcal{B}_2(x-t)\,dt = \cdots = \frac{1}{\pi}\int_{\mathbb{T}} f^{(m)}(t)\mathcal{B}_m(x-t)\,dt\,. \end{aligned} \tag{5.15}$$

Conversely, let f have the representation (5.14) with some $\phi \in L_p(\mathbb{T})$. By what we know, for the r-th periodic integral Φ of ϕ, $\Phi = \phi * \mathcal{B}_m$. We have $\Phi = f$ and therefore $\phi = f^{(r)}$, a.e.

$$\Phi^{(r)} = f^{(r)}\,,\ \text{a.e.}\,. \qquad \square$$

The $\mathcal{E}_m$ are examples of *periodic perfect splines,* while the $\mathcal{B}_m$ are multiples of *monosplines* (see §1).

§6. Definition of Splines by Their Extremal Properties

The word "spline" comes from an instrument called a mechanical spline which was used years ago by draftsmen as a means of producing a smooth curve through a given finite set $P := \{(x_i, y_i)\}_1^n$, $a < x_1 < \cdots < x_n < b$, of planar points. The points were represented by positions on the draftsman's board and the curve by a thin pliable rod. Metal ducks were used to position the rod through the prescribed points and then the final configuration of it was the desired curve.

It turns out that the rod will occupy a position which minimizes its strain energy. The strain energy in first approximation is the integral $\int_I (f'')^2 dx$ with

$I := [a,b]$. Thus, the draftsman's problem is to minimize this integral for all curves $y = f(x)$ that satisfy the interpolation conditions

$$f(x_i) = y_i\,, \quad i = 1,\dots,n\,. \tag{6.1}$$

We shall see that the solution to this problem is a *cubic spline function.* In fact, we shall solve the following more general extremal problem for each integer $m < n$:

(*) *Among all admissible functions $f \in W_2^m$ (that is, functions which satisfy (6.1) and have a derivative $f^{(m)} \in L_2$), one has to find functions which minimize the integral*

$$\int_I \left(f^{(m)}\right)^2 dx\,. \tag{6.2}$$

To $x_1,\dots,x_n$, we adjoin points $x_j \le a$, $j = -2m+1,\dots,0$ and $x_j \ge b$, $j = n+1,\dots,n+2m$ in an arbitrary manner. For the resulting sequence X, we let $\mathcal{S}^0$ be the space of all splines S in $\mathcal{S}_{2m}(X,I)$ that on (a,x_1) and on (x_n,b) are polynomials of degree $< m$. Splines of this type are sometimes called *natural splines.*

The splines $S \in \mathcal{S}^0$ are identical with m-th integrals of splines $T \in \mathcal{N}_m$, the space $\mathcal{N}_m := \mathcal{N}_m(X,I)$ consisting of all T which are linear combinations of the B-splines $M_j(x) := M(x;x_j,\dots,x_{j+m})$, $j = 1,\dots,n-m$:

$$T = \sum_{j=1}^{n-m} a_j M_j\,. \tag{6.3}$$

Indeed, from the differentiation formula (3.11), we have $S^{(m)} = \sum_{-m+1}^{n} a_j M_j$ on I. The M_j, $j = -m+1,\dots,0$ and $j = n-m+1,\dots,n$ are linearly independent on (a,x_1) and (x_n,b), respectively. Hence $S^{(m)}$ will vanish on these intervals, exactly when $a_j = 0$ for these values of j. We obtain therefore (6.3).

The following theorem shows that the problem (*) has a spline as its unique solution.

Theorem 6.1 (de Boor [1963]), Schoenberg [1964]). *The extremal problem (*) has a unique solution; this solution is the unique admissible spline from $\mathcal{S}^0$.*

Proof. Let $f \in W_2^m$ be an arbitrary admissible function. We approximate $f^{(m)}$ in the L_2 metric by elements of $\mathcal{N}_m$. If T is the best approximation, then

$$\int_I (f^{(m)} - T) M_j dx = 0\,, \quad j = 1,\dots,n-m\,. \tag{6.4}$$

$$\int_I T^2 dx \le \int_I (f^{(m)})^2 dx\,. \tag{6.5}$$

Let S_0 be one of the m-th integrals of T. We can select S_0 to satisfy

$$S_0(x_i) = f(x_i) = y_i\,, \quad i = 1,\dots,m\,. \tag{6.6}$$

On the other hand, from (2.2), we can write (6.4) as

$$[x_j,\dots,x_{j+m}](f - S_0) = 0\,, \quad j = 1,\dots,n-m\,. \tag{6.7}$$

We can use formula (7.7) of Chapter 4 and obtain by induction $f(x_i) - S_0(x_i) = 0$, $i = 1,\dots,n$. Thus $S_0 \in \mathcal{S}^0$ is admissible.

There is only one admissible element in $\mathcal{S}^0$. For if S_1 is another one, then $(S_1 - S_0)(x_i) = 0$ for all i implies $[x_j,\dots,x_{j+m}](S_1 - S_0) = 0$ or

$$\int_I \left(S_1^{(m)} - S_0^{(m)}\right) M_j dx = 0\,, \quad j = 1,\dots,n-m\,. \tag{6.8}$$

Since $S_1^{(m)} - S_0^{(m)}$ is in $\mathcal{N}_m$, this means $S_1^{(m)} - S_0^{(m)} \equiv 0$. It follows that $S_1 - S_0$ is a polynomial of degree $< m$ which vanishes at $x_1,\dots,x_m$; hence $S_1 \equiv S_0$.

Since $S_0^{(m)} = T$, we have $\int_I [S_0^{(m)}]^2 dx \le \int_I [f^{(m)}]^2 dx$ for all admissible f. Therefore S_0 is a solution to (*). If g is any other solution, $\int_I [g^{(m)}]^2 dx = \int_I [S_0^{(m)}]^2 dx$, then $\|g^{(m)} - S_0^{(m)}\|_2 = \|g^{(m)}\|_2 - \|S_0^{(m)}\|_2 = 0$, hence $g^{(m)} = S_0^{(m)}$, a.e., and integrating with a view to (6.1), we obtain $g = S_0$. □

It is useful to note that Theorem 6.1 holds (with an identical proof) for a bi-infinite knot sequences $T := (t_i)$ and bi-nifinite sequences $X := (x_i)$ of interpolation points and data $Y := (y_i)$ provided there is an f with $f^{(m)} \in L_2(\mathbb{R})$ which satisfies $f(x_i) = y_i$, $i \in \mathbb{Z}$.

§ 7. The Kolmogorov-Landau Inequalities

In Chapter 2, we have shown that for any subinterval I of $\mathbb{R}$ and for $f \in W_\infty^m(I)$, $m \ge 2$, we have with all norms over I,

$$\|f^{(k)}\|_\infty \le C\|f\|_\infty^{1-k/m}\|f^{(m)}\|_\infty^{k/m}\,, \quad k = 0,\dots,m\,; \tag{7.1}$$

the constant C depends on k, m and I. Remarkably, for certain intervals I, one can find the best constants in (7.1). The first result of this type was Landau's [1913] determination of the constant $C = \sqrt{2}$ for $m = 2$, $k = 1$ and $I = \mathbb{R}$. Later, Kolmogorov found the best constants in (7.1) for $\mathbb{R}$, for all m and k. Let

$$C_{m,k} := K_{m-k} K_m^{-1+k/m}\,, \quad k = 0,1,\dots,m\,, \tag{7.2}$$

where K_j are the Favard constants (5.5). Kolmogorov [1939] proved

Theorem 7.1. *For $m = 2,3,\dots,$ and $0 \le k \le m$, we have for any $f \in W_\infty^m(\mathbb{R})$,*

$$\|f^{(k)}\|_\infty \le C_{m,k}\|f\|_\infty^{1-k/m}\|f^{(m)}\|_\infty^{k/m}\,. \tag{7.3}$$

Moreover, the constant in (7.3) *cannot be replaced by a smaller constant: there is equality in* (7.3) *for all dilated Euler splines* $\mathcal{E}_m(x/a)$, *for* $a > 0$. *The same statement applies to the space* $W^m_\infty(\mathbb{T})$.

There is an alternate formulation of this theorem.

Theorem 7.2. *For* $F \in W^m_\infty$, *the inequalities*

$$\|F\|_\infty \le \|\mathcal{E}_m\|_\infty = 1 \text{ and } \|F^{(m)}\|_\infty \le \|\mathcal{E}_m^{(m)}\|_\infty = \pi^m/K_m \tag{7.4}$$

imply that

$$\|F^{(k)}\|_\infty \le \|\mathcal{E}_m^{(k)}\|_\infty, \quad k = 1, \dots, m-1. \tag{7.5}$$

Indeed, this is a special case of (7.3) (see (5.7)). On the other hand, we can also show that Theorem 7.2 implies (7.3). For an arbitrary $f \in W^m_\infty$, we can assume that $\|f\| =: M$ and $\|f^{(m)}\| =: M_1$ are not zero. We put $F(x) := \alpha f(\beta x)$, $\alpha := M^{-1}$, $\beta := (M\pi^m M_1^{-1} K_m^{-1})^{1/m}$ and from (7.5) obtain

$$\|f^{(k)}\|_\infty = \alpha^{-1}\beta^{-k}\|F^{(k)}\|_\infty \le \alpha^{-1}\beta^{-k}\|\mathcal{E}_m^{(k)}\|_\infty,$$

which is (7.3).

In our proof of Theorem 7.2, we shall follow Cavaretta [1974] which is simple if one uses some elementary refinements of Rolle's theorem. We begin by establishing this theorem for periodic functions.

Lemma 7.3. *The statement of Theorem 7.2 is true for* $2n$ *periodic functions* $F \in C^{(m)}(\mathbb{R})$, *for each* $n = 1, 2, \dots$.

We restrict such functions to the interval $I := [-n, n]$ with endpoints identified. All Euler splines $\mathcal{E}_k$ are $2n$ periodic and have exactly $2n$ sign changes on I.

We shall need the following forms of Rolle's theorem in which f is a function on $\mathbb{T}$; they are easy to prove by examining f' between two zeros of f. *We assume that* f' *does not vanish identically on any interval.*

(I) If a continuously differentiable function f has p distinct zeros on $\mathbb{T}$, then f' has p distinct zeros on $\mathbb{T}$, each with a sign change.

(Ia) If in addition, f and f' vanish at a point x_0 (which may be one of the zeros of (I)), then f' has at least $p+1$ distinct zeros.

(II) If f is continuous and piecewise differentiable and has p distinct zeros on $\mathbb{T}$, then there are p disjoint intervals on which f' changes sign.

Proof of Lemma 7.3. If the Lemma were not true, we could find a $2n$ periodic function F in $C^m(\mathbb{R})$ which satisfies (7.4) and in addition for some k, $1 \le k < m$ and $q < 1$, for the uniform norm on $\mathbb{R}$, has the property

$$q\|F^{(k)}\| = \|\mathcal{E}_m^{(k)}\|. \tag{7.6}$$

The idea of the proof is as follows. We examine the $2n$ periodic function

$$G(x) := \mathcal{E}_m(x) - qF(x),$$

on $\mathbb{R}$ with points $x \in \mathbb{R}$ *identified modulo* $2n$. We show that it has $2n$ zeros. From this we derive that $G^{(m)}$ has $2n+1$ changes of sign, and this leads to a contradiction.

With $F(x)$, also $\pm F(x+c)$ satisfies (7.4) and (7.6), so we can assume that $\mathcal{E}_m^{(k)}(x_0) = \|\mathcal{E}_m^{(k)}\|$ and $qF^{(k)}(x_0) = \|qF^{(k)}\|$ at some point x_0. The function G is in $C^{(m-1)}$ and $G^{(m-1)}$ is piecewise differentiable. From $\|qF\|_\infty < \|\mathcal{E}_m\|_\infty$, we derive that G has $2n$ extrema of opposite sign, hence it has $2n$ zeros. Moreover, $G^{(m)}$ is not identically zero on any interval and therefore the same is true for $G^{(j)}$, $j = 0, \dots, m-1$.

If $k < m-1$, we use (I) to see that each $G^{(j)}$, $1 \le j \le k$ has at least $2n$ zeros. In addition, $G^{(k)}$ has a zero of multiplicity at least two at x_0. Then (Ia) and (I) yield that for $k < j \le m-1$, $G^{(j)}$ has $2n+1$ zeros. Now let $k = m-1$. From (I) we get that $G^{(j)}$, $j \le m-1$, has at least $2n$ zeros (which are sign changes). The point x_0 is also a zero of $G^{(m-1)}$. The one-sided derivatives of $\mathcal{E}_m^{(m-1)}$ at x_0 are $\pm c$, $c > 0$, while the derivative of $qF^{(m-1)}$ vanishes at x_0. Hence $G^{(m-1)}$ does not change sign at x_0. We have found an additional zero of $G^{(m-1)}$.

In both cases, we have shown that $G^{(m-1)}$ has $2n+1$ distinct zeros and between any pair of these, it is not identically zero. By (II), there are $2n+1$ disjoint intervals where $G^{(m)}$ changes sign. Since $q < 1$, by (7.4), $q\|F^{(m)}\| < \|\mathcal{E}_m^{(m)}\|$, the function $g^{(m)}$ changes sign only when the step function $\mathcal{E}_m^{(m)}$ does. Since this happens only $2n$ times, we have a contradiction. □

Proof of Theorem 7.2. We show first that (7.5) holds when F has compact support. For this, we observe that for each $\varepsilon > 0$, there is a $G \in C^m$ of compact support satisfying: $\|G^{(m)}\| \le \|F^{(m)}\|$ and $\|F^{(k)} - G^{(k)}\| < \varepsilon$, $k = 0, 1, \dots, m-1$. For example, if $K_\eta(x) := \eta^{-1}K(t/\eta)$ where K is a non-negative, compactly supported function in $C^{(m)}$ of integral one, then $G(x) := (F * K_\eta)(x) = \int F(y)K_\eta(x-y)\,dy$ has these properties whenever $\eta > 0$ is sufficiently small. We can therefore assume additionally that F is in $C^m(\mathbb{R})$. Now if n is so large that F vanishes outside of $[-n, n]$ and on some neighborhoods of the endpoints, then the $2n$-periodic extension of F will have period $2n$ and satisfy (7.4). By the Lemma, this new function satisfies (7.5) and therefore so does F.

Finally, let $F \in W^m_\infty(\mathbb{R})$ be an arbitrary function which satisfies (7.4). If $A \ge 1$ is arbitrary, we let ϕ be a compactly supported function of norm one which is one on $[-A, A]$ and satisfies $\|\phi^{(i)}\| \le 1/A$, $i = 1, \dots, m$, (it is easy to construct such ϕ by integration). For any $\lambda < 1$, the function $H := \lambda\phi F$ satisfies $\|H\| \le 1$ and by Leibnitz's rule

$$\|H^{(m)}\| \le \lambda \sum_{i=0}^m \binom{m}{i} \|\phi^{(m-i)}\|\,\|F^{(i)}\| \le \lambda\left(\|F^{(m)}\| + \frac{2^m}{A}\sum_{i=1}^m \|F^{(i)}\|\right).$$

Hence, if A is sufficiently large, H satisfies (7.4) and has compact support. By what we have already proved, for $|x| \le A$,

$$(7.7) \qquad \lambda \left|F^{(k)}(x)\right| \le \left\|H^{(k)}\right\| \le \left\|\mathcal{E}_m^{(k)}\right\|.$$

If we now let $A \to \infty$ and then $\lambda \to 1$, we obtain (7.5). □

For the function $\mathcal{E}_m(x)$ and also for any $A\mathcal{E}_m(ax+b)$, we have equality in (7.3). Some of the values of $C_{m,k}$ are

$$(7.8) \qquad C_{2,1} = \sqrt{2}\,, \quad C_{3,1} = \frac{1}{2}\sqrt[3]{9}\,, \quad C_{3,2} = \sqrt[3]{3}\,.$$

Inequality (7.3) is valid in other functions spaces, although the best constants are not always known.

Theorem 7.4 (Stein [1957]). *With the constant $C_{m,k}$ of (7.2), one has for $f \in W_p^m(\mathbb{R})$, $1 \le p \le \infty$,*

$$(7.9) \qquad \left\|f^{(k)}\right\|_p \le C_{m,k} \left\|f\right\|_p^{1-k/m} \left\|f^{(m)}\right\|_p^{k/m}, \quad k = 1, \dots, m-1\,.$$

The same is true for functions $f \in W_p^r(\mathbb{T})$.

Proof. We prove this theorem for the case of $\mathbb{R}$; a similar proof applies for $\mathbb{T}$. Let $g \in L_{p'}$, $1/p + 1/p' = 1$. The function

$$(7.10) \qquad F(x) := \int_{-\infty}^{\infty} f(x+t)g(t)dt$$

belongs to $W_\infty^m(\mathbb{R})$ and has derivatives

$$F^{(k)}(x) = \int_{-\infty}^{\infty} f^{(k)}(x+t)g(t)dt\,, \quad k = 1, \dots, m\,.$$

For a fixed k, $0 < k < m$, we put

$$(7.11) \qquad g(x) := \begin{cases} \operatorname{sign} f^{(k)}(x)\,, & \text{if } p = 1\,, \\ \dfrac{\left|f^{(k)}(x)\right|^{p-1} \operatorname{sign} f^{(k)}(x)}{\left\|f^{(k)}\right\|_p^{p/p'}}\,, & \text{if } 1 < p < \infty\,. \end{cases}$$

Then $\|g\|_{p'} = 1$ and

$$F^{(k)}(0) = \int_{-\infty}^{\infty} f^{(k)}(t)g(t)dt = \left\|f^{(k)}\right\|_p.$$

To the function F, we apply Kolmogorov's inequality (7.3) and obtain

$$\left|F^{(k)}(0)\right| \le C_{m,k} \|F\|_\infty^{1-k/m} \left\|F^{(m)}\right\|_\infty^{k/m}.$$

Since $\|F\|_\infty \le \|f\|_p \|g\|_{p'} = \|f\|_p$ and similarly, $\left\|F^{(m)}\right\|_\infty \le \left\|f^{(m)}\right\|_p$, we obtain (7.9). □

When $p = 1$, it is actually enough to assume that $f^{(m-1)} \in BV$. Then (7.9) becomes

$$(7.12) \qquad \left\|f^{(k)}\right\|_1 \le C_{m,k} \|f\|_1^{1-k/m} \left[\operatorname{Var}\left(f^{(m-1)}\right)\right]^{k/m}.$$

This is proved by approximating f by smooth functions g. For example in the case of $\mathbb{R}$ and f of compact support, there is a sequence g_n of functions in $W_1^m(\mathbb{R})$ such that $g_n^{(k)}$ converges to $f^{(k)}$ uniformly as $n \to \infty$ and $\|g_n^{(m)}\|_1 \le \operatorname{Var}(f^{(m-1)})$. For example, we can take $g_n := f * K_{\eta_n}$ with kernels K_η as in the proof of Theorem 7.2. We apply (7.9) to the g_n and then take a limit as $n \to \infty$ to obtain (7.12).

The constant $C_{m,k}$ of (7.12) is best possible. For example, on $\mathbb{T}$, $f(x) := g'(x)$ where $g(x) := \mathcal{E}_m(x/\pi)$ gives equality in (7.12). Indeed, $\|g^{(j)}\|_1 = \operatorname{Var}(g^{(j-1)}) = 4\|g^{(j-1)}\|_\infty = 4\pi^{-j+1}\|\mathcal{E}_m^{(j-1)}\|_\infty$. Hence, we need only use the fact that $\mathcal{E}_m$ gives equality in (7.1). The constant $C_{m,k}$ is also best possible in (7.9) for the case of $\mathbb{R}$ (but it is not attained). This can be derived from the periodic case using ideas similar to the proof of Theorem 7.2.

§ 8. Zero Count for Splines

We derive here upper estimates for the number of zeros of a spline function S. The situation is complicated by two facts. First, S (and its derivatives) may vanish on intervals; we call these *zero intervals*. Second, derivatives of S need not be continuous. To deal, in part, with the first of these difficulties, we shall count the zeros of S only on its *support intervals*. These are maximal intervals I such that S does not vanish identically on any subinterval of I.

Let $I = [a,b]$ be a support interval of S and $c \in (a,b)$. The product $S(c-h_1)S(c+h_2)$ is either always < 0 or always > 0 for all small $h_1, h_2 > 0$. In the first case S *changes sign* at c, in the second it *does not*. We define the *point zeros* $c \in I$ of S and their *multiplicities* as follows. A point of discontinuity $c \in (a,b)$ of S is a *discontinuous zero* of S and its multiplicity is 1 if and only if S changes sign at c. In particular a discontinuous zero can not occur at the endpoint of a support interval. *Continuous zeros* $c \in I$ are defined in the usual way. Their multiplicity is the largest integer l such that c is a continuous zero of $S, \dots, S^{(l-2)}$ and a continuous or discontinuous zero of $S^{(l-1)}$.

Here is an important observation that we shall use frequently in our zero count. If $c \in (a,b)$ is a zero of S of multiplicity $l+1$, then c cannot be in a zero interval of $S^{(l)}$. Indeed, if this were not the case, then $S^{(l)}$ and therefore also S would vanish in some interval which contains c and this contradicts the definition of the support intervals of S.

We let $Z := Z(S, (a, b))$ denote the number of zeros of S in I, counting multiplicity. The main result of this section is Theorem 8.2 which gives an upper estimate for Z. A useful tool for counting spline zeros is the following Rolle theorem (Lorentz [1975]):

Theorem 8.1. *Let S be a spline such that S, S' do not vanish identically on any subinterval of (c_1, c_2). If $S(c_1+) = S(c_2-) = 0$, then for some point ξ, with $c_1 < \xi < c_2$, either* (A)*: ξ is a zero of S' with change of sign, but not a zero of S, or,* (B)*: ξ is a discontinuous zero of S, at which S' preserves sign (hence, not a zero of S').*

Proof. We can assume that (c_1, c_2) contains no other continuous zeros of S, for otherwise we could replace the interval by a smaller one. For all small $h > 0$, $S(c_1 + h)S'(c_1 + h) > 0$ and $S(c_2 - h)S'(c_2 - h) < 0$. Therefore, S and S' have in (c_1, c_2) numbers of sign changes of different parity. Hence, in transversing (c_1, c_2) from left to right, there exists a first point ξ at which S' but not S, or S but not S' changes sign. In the first case, ξ cannot be a zero of S (continuous or discontinuous) and we have (A); in the second case, ξ is a discontinuous zero of S as in (B). □

We emphasize that the point ξ of Theorem 8.1 cannot be a zero of both S and S'.

Let S be a spline function of degree m on one of its support interval $I = [a, b]$ and let $a =: t_0^* < t_1^* < \cdots < t_n^* := b$ be its breakpoints in I. We use the generic notation I_j to denote the support intervals of $S^{(j)}$, $j = 0, 1, \ldots, m$. The endpoints of I_j are breakpoints and each I_j is contained in one of the I_{j-1}. We associate to S the "diagram" $\Phi := \Phi_S$ which is the set of all points $(x, y) \in \mathbb{R}^2$ such that x is in one of the I_j and $y \le j$. Hence, Φ is the union of the (not necessarily disjoint) rectangles $I_j \times [0, j]$. For $j = 0, 1, \ldots, m$, we let l_j be the number of the intervals I_j and we put

$$L(\Phi) := \sum_1^m l_j \,. \tag{8.1}$$

It is easy to see that $2L(\Phi)$ is the length of the vertical boundary of Φ.

A pair (t_i^*, j) is said to be *singular* for $S^{(j)}$ if t_i^* is in a support interval of $S^{(j)}$ and one of the functions $S, \ldots, S^{(j)}$ is discontinuous at t_i^*. For example, if S is given by (1.6), then (t_i^*, j) is singular if and only if $r - k_i \le j \le r - 1$ and $c_{i,k} \ne 0$ for some $k \ge j$. It follows that if (t_i^*, j) is a singular pair then so is (t_i^*, k) for all $k > j$. All singular pairs are contained in Φ. We write σ_j for the number of singular pairs of $S^{(j)}$ contained in I, and put $\sigma := \sum_0^m \sigma_j$.

Theorem 8.2 (Jetter, Lorentz, Riemenschneider [1983]). *If S is a spline function and $I = [a, b]$ is one of its support intervals, then*

$$Z := Z(S, (a, b)) \le \sigma - L(\Phi) - 2 \,. \tag{8.2}$$

Proof. Multiple zeros of S are also zeros of some of the derivatives $S^{(j)}$. We count Z in the following way. Let z_j be the number (not counting multiplicity) of *continuous* zeros of $S^{(j)}$ in the interior of one of the intervals I_j of level j, which are zeros of S of multiplicity $\ge j+1$ (hence also zeros of $S, \ldots, S^{(j-1)}$). Let z_j^* be the number of *discontinuous* zeros of $S^{(j)}$ in the interior of the same intervals of multiplicity $j + 1$ (hence not a zero of $S^{(j+1)}$ if $j < m$). By a previous remark, a zero of S of multiplicity $j + 1$ counted in Z cannot be an endpoint of an I_j and therefore

$$Z = \sum_{j=0}^m (z_j + z_j^*) \,. \tag{8.3}$$

Further, let σ_j' and σ_j'' be the numbers of singular pairs (t_i^*, j) in the interior of the I_j and at the endpoints of the I_j, respectively. Then

$$\sigma = \sum_{j=0}^m (\sigma_j' + \sigma_j'') \,. \tag{8.4}$$

The proof consists in applying Rolle's Theorem 8.1 to obtain new (that is, different from those counted in z_j) continuous point zeros of $S^{(j)}$. We let μ_j be a lower estimate of the number of new continuous zeros of $S^{(j)}$ obtained in this way. We begin with $\mu_0 = 0$ and show how to estimate μ_{j+1} from μ_j.

In addition to the continuous zeros of $S^{(j)}$ counted in z_j and μ_j, there are also the zeros at the endpoints t_i^* of the I_j at which $S^{(j)}$ is continuous. There are $2l_j - \sigma_j''$ of these. Hence, there are at least $\mu_j + z_j + 2l_j - \sigma_j''$ continuous zeros of $S^{(j)}$, each of them contained in one of the support intervals I_j. Also each of these zeros is contained in a support interval I_{j+1} of $S^{(j+1)}$ since otherwise $S^{(j)}$ would be identically zero in a neighborhood of this zero. We want to apply Rolle's theorem to intervals U formed by consecutive pairs of these zeros but we must exclude those pairs whose zeros lie in different I_{j+1}. There are at most $l_{j+1} - 1$ of such pairs. There remains $\mu_j + z_j + 2l_j - \sigma_j'' - l_{j+1}$ intervals U. We apply Theorem 8.1 to these U and obtain that many points ξ. Some of the ξ will be new continuous zeros of $S^{(j+1)}$ (special case of (A)). Any other possibility would require that $\xi = t_i^*$ with t_i^* in the interior of U. Now, either ξ is a discontinuous zero of $S^{(j)}$ or a discontinuous zero of $S^{(j+1)}$ but not both. Hence $(\xi, j + 1)$ is a singular point of $S^{(j+1)}$. Also ξ is *not* one of the points counted in z_{j+1}^* because they are all continuous zeros of $S^{(j)}$. There are therefore at most $\sigma_{j+1}' - z_{j+1}^*$ such ξ. For $j = 0, \ldots, m-1$, we obtain $\mu_{j+1} \ge \mu_j + z_j + 2l_j - l_{j+1} - \sigma_j'' - \sigma_{j+1}' + z_{j+1}^*$. We move μ_j to the left side and sum over $j = 0, \ldots, m - 1$ to obtain

$$0 = \mu_m \ge \sum_0^{m-1} z_j + 2\sum_0^{m-1} l_j - \sum_1^m l_j - \sum_0^{m-1} \sigma_j'' - \sum_1^m \sigma_j' + \sum_1^m z_j^* \,.$$

Since $z_m = 0$, $l_0 = 1$, $2l_m = \sigma_m''$, we get

$$0 \geq \sum_0^m (z_j + z_j^*) - z_0^* + \sum_1^m l_j + 2 - \sum_0^m (\sigma_j' + \sigma_j'') + \sigma_0' .$$

The first sum is $Z(S,(a,b))$, the second sum is $L(\Phi)$, the third sum is σ, and $z_0^* \leq \sigma_0'$. We thus obtain (8.2). □

In particular, (8.2) can be applied to splines in a Schoenberg space, $\mathcal{S}_r := \mathcal{S}_r(T,A)$, or more generally to any spline S which is a linear combination of B-splines. If $T = (t_i)$ is a knot sequence (finite or infinite), we let $\mathcal{S}_r(T)$ denote the space spanned by the B-splines N_j of order r for T. Splines S in $\mathcal{S}_r(T)$ are defined on all of $\mathbb{R}$. The number of singular pairs $\sigma = \sigma(I)$ for S in an interval $I = [c,d]$ cannot exceed the number $k(I)$ of knots (counting multiplicity) in I. Indeed, if the knot t_i is repeated k_i times in T then each B-spline N_j has $r - k_i - 1$ continuous derivatives at t_i and therefore the same is true for each $S \in \mathcal{S}_r(T)$.

Corollary 8.3. *If $S \in \mathcal{S}_r(T)$ and $I = [c,d]$ is a support interval of S, then*

$$Z(S,(c,d)) \leq k(I) - r - 1 . \tag{8.5}$$

If I is an arbitrary interval outside of which S vanishes and if I contains s support intervals of S, then

$$Z(S,(c,d)) \leq k(I) - s(r+1) . \tag{8.6}$$

Proof. On any support interval $L(\Phi) \geq r-1$ and therefore (8.5) follows from (8.2). Then (8.6) follows by applying (8.5) on each of its support intervals. □

It follows from (8.5) that a support interval I of S must contain at least $r+1$ knots. We have given an independent proof of this fact earlier using the de Boor-Fix formula. We can also use (8.5) to give a new *proof of the Curry-Schoenberg Theorem* 3.1 that the N_j, $j = -r+1, \ldots, n$, are a basis for $\mathcal{S}_r(T,A)$, $A = [a,b]$. As before it is enough to show that the B-splines $(N_j)_{-r+1}^n$ are linearly independent on A. If $S_0 := \sum_{-r+1}^n a_j N_j \equiv 0$ on A, then S_0 vanishes outside of $I := [t_{-r+1}, t_0]$. But I contains only r knots and therefore S_0 vanishes identically on I and hence for all $t \leq a$. Similarly, S_0 vanishes for $t \geq b$. Now let k be the multiplicity of the knot $\tau := t_{-r+1}$. Among the splines N_j only those with $j = -r+1, \ldots, -r+k$ are different from zero in each right neighborhood of τ. By (2.6), we have $N_{-r+k}(t) = (t-\tau)^{r-k}$ in this neighborhood. It follows that $a_{-r+1} = = \cdots = a_{-r+k} = 0$. Continuing in this way, we get $a_{-r+1} = \cdots = a_n = 0$.

As another interesting application of Corollary 8.3, we prove, as anounced in §1:

Theorem 8.4. *The Schoenberg space $\mathcal{S}_r(T,A)$ is a weak Haar space. More generally, if $T = (t_i)_1^{n+r}$, then $\mathcal{S}_r(T)$ is a weak Haar space on $\mathbb{R}$.*

Proof. Let $S \in \mathcal{S}_r(T) = \operatorname{span}(N_j)_1^n$. If $S(c)S(d) < 0$, then the interval $[c,d]$ contains either an interval zero of S or a (continuous or discontinuous) point zero of S. There are at most $s-1$ such interval zeros where s is the number of support intervals of S in $\mathbb{R}$. Since there are a total of $n+r$ knots, (8.6) gives $Z(S,(a,b)) \leq n+r-s(r+1)$. Therefore S has at most $n+r-s(r+1)+s-1 \leq n-1$ sign changes. □

§9. Spline Interpolation

We shall discuss here Lagrange and Hermite interpolation by splines. In contrast to polynomial interpolation, these problems are not always solvable. Certain relations between the knots of splines and the interpolation points must be satisfied.

We shall show that these problems allow a simple solution by means of the zero count theorems of §8. We begin by discussing interpolation on $\mathbb{R}$.

Let $T := (t_i)_1^{n+r}$ be a finite knot sequence (with possible repetitions), let $\mathcal{S}_r(T) := \mathcal{S}_r(T,\mathbb{R})$ be the space spanned by B-splines $N_j(x) = N(x; t_j, \ldots, t_{j+r})$, $j = 1, \ldots n$, of order r. Its dimension is n; we consider a sequence of n interpolation points $X : x_1 \leq \cdots \leq x_n$ on $\mathbb{R}$. As in §6 of Chapter 4, repetition of a point x_i means the prescription of some derivative values of $S \in \mathcal{S}_r(T)$ at x_i. More precisely, let l_i, $i = 1, \ldots, n$, signify the number of x_j with $x_j = x_i$, $j < i$. The Hermite interpolation problem for $\mathcal{S}_r(T)$ with data (y_i) consists in finding an $S \in \mathcal{S}_r(T)$ for which

$$S^{(l_i)}(x_i) = y_i, \quad i = 1, \ldots, n . \tag{9.1}$$

(If all x_i, are distinct, this is a Lagrange interpolation problem.)

However, we want to avoid the case when $S^{(l_i)}$ is not defined at x_i. For this purpose, we assume that if $x_i = t_j$, then $l_i - 1$ is less than $r - k_j$, which is the order of the first derivative of $S \in \mathcal{S}_r(T)$ that may not exist at t_j. Thus, we shall consider only problems (9.1) for which

$$l_i + k_j \leq r , \quad \text{if } x_i = t_j , \tag{9.2}$$

or, in other words, problems such that any point x appears a total of at most r times in X and T.

Our problem is not always solvable. For example, if all points x_i lie in a single interval (t_j, t_{j+1}), then there are too many conditions imposed on the polynomial piece of S in this interval to allow a solution.

Theorem 9.1. *For given X and T from $\mathbb{R}$ with the property* (9.2)*, the problem* (9.1) *is uniquely solvable if and only if X and T satisfy the interlacing condition*

$$t_i < x_i < t_{i+r} , \quad i = 1, \ldots, n . \tag{9.3}$$

Proof. The problem (9.1) is solvable if and only if the homogeneous problem with $y_i = 0$ has only the trivial solution. To see that (9.3) is necessary, we assume that $x_i \le t_i$ for some i, and consider the splines $S_0 = \sum_{j=i}^n c_j N_j$. They and all their derivatives that appear in conditions (9.1) with $j \le i$, vanish. There are only $n - i$ remaining conditions (9.1) with $j > i$, and $n - i + 1$ free coefficients c_i. Hence there is a non-trivial S_0 which vanishes also at the x_j, $j > i$. A similar argument applies when $x_i \ge t_{i+r}$ for some i.

To show that the conditions are sufficient, let $S \in \mathcal{S}_r(T)$ be a non-trivial spline which satisfies the homogeneous equations (9.1). Let $I := [t_i, t_{i+s}]$, $s \ge r$ be one of its support intervals, as defined in §8. By (9.3), I contains $s - r + 1$ (point) zeros of S in its interior. But I contains only $s + 1$ knots. This contradicts (8.5). $\square$

We can give another formulation for Theorem 9.1. For the B-spline basis $N_1, \dots, N_n$ of $\mathcal{S}_r(T)$, the matrix of the equations (9.1) is the *collocation matrix*

(9.4) $$C^* := C^*(X, T) = \left(N_j^{(l_i)}(x_i)\right)_{i,j=1}^n .$$

We have proved that C^* is *non-singular if and only if X and T satisfy* (9.3).

Almost identical are the interpolation problems for a Schoenberg space $\mathcal{S}_r(T, I)$, $T = (t_i)_{-r+1}^{n+r}$ on an interval $I = [a, b]$ instead of $\mathcal{S}_r(T, \mathbb{R})$. The only difference is that the knots t_i, $i = -r+1, \dots, 0$ and $i = n+1, \dots, n+r$ are outside (a, b), while the $n + r$ interpolation points x_i are contained in I:

(9.5) $$a \le x_{-r+1} \le \cdots \le x_n \le b .$$

Since the B-splines N_j, $j = -r+1, \dots, n$, are a basis for $\mathcal{S}_r(T, I)$ we obtain the following theorem, due to Schoenberg and Whitney [1953] for Lagrange interpolation, and to Karlin and Ziegler [1966] for the Hermitian case.

Theorem 9.2. *The interpolation problem* (9.1) *for $i = -r+1, \dots, n$ is uniquely solvable for some $S \in \mathcal{S}_r(T, I)$ for the interpolation points* (9.5) *if an only when X and T satisfy the interlacing condition* (9.3).

However, some of the auxiliary knots t_j, $j \le 0$ and $j > n$ may be outside of I, then some of the inequalities (9.3) are automatically satisfied. We can omit the first inequality in (9.3) when $t_i < a$, and the second inequality when $t_{i+r} > b$.

§ 10. Sign Variation of Splines

The representation of a spline S in its B-spline series carries much information about the shape of the graph of S. For example, S is non-negative whenever its B-spline coefficients c_j are all non-negative. More generally, the B-spline series has the *sign variation diminishing* property. That is, the spline S will change sign on $(-\infty, \infty)$ no more often than the sequence of coefficients $(c_j)_{-\infty}^{\infty}$ changes sign.

We confine the proof to splines on $\mathbb{R}$ with bi-infinite $T := (t_i)_{-\infty}^{\infty}$; statements about splines on a compact interval follow from this automatically. We follow Lane and Riesenfeld [1983] and consider knot sequences U which are *refinements* of T, that is, sequences U which contain T as a subsequence. If $S \in \mathcal{S}_r(T) := \mathcal{S}_r(T, \mathbb{R})$, then S is also in $\mathcal{S}_r(U)$ and therefore can be represented by two B-spline series:

(10.1) $$S = \sum_{j \in Z} c_{j,T} N_{j,T} = \sum_{j \in Z} c_{j,U} N_{j,U} ,$$

where the $N_{j,T}$ and $N_{j,U}$ are the B-splines for T and U respectively.

We want to find relations between the coefficients in these two series. We begin with the simplest case where U is a one knot refinement of T. This means that U is of the form $U : \dots, t_{k-2}, t_{k-1}, u_k, t_k, t_{k+1}, \dots$ for some additional knot u_k. We would like to compare the difference operators $T_j := [t_j, \dots, t_{j+r}]$ and $U_j := [u_j, \dots, u_{j+r}]$. In the case $j \ge k$, all t_i appearing in T_j satisfy $t_i = u_{i+1}$, hence $T_j = U_{j+1}$; in the case $j < k - r$, they all satisfy $t_i = u_i$, hence $T_j = U_j$.

In the remaining case, we have $u_j = t_j$, $u_{j+r+1} = t_{j+r}$ and from (7.5) of Chapter 4, we obtain

$$\frac{U_{j+1} - U_j}{u_{j+r+1} - u_j} = [u_j, \dots, u_{j+r+1}] = \frac{T_j - U_j}{u_{j+r+1} - u_k} , \quad k - r \le j < k .$$

Therefore,

(10.2) $$T_j = \frac{u_{j+r+1} - u_k}{u_{j+r+1} - u_j} U_{j+1} + \left(1 - \frac{u_{j+r+1} - u_k}{u_{j+r+1} - u_j}\right) U_j , \quad k - r \le j < k .$$

This gives a corresponding identity for B-splines. We define

$$\alpha_i := \min\left(\frac{(u_{i+r} - u_k)_+}{u_{i+r} - u_i}, 1\right) ,$$

$i \in \mathbb{Z}$. Then from the definition (2.7) of the B-splines, we obtain

(10.3) $$N_{j,T} = \alpha_{j+1} N_{j+1,U} + (1 - \alpha_j) N_{j,U} , \quad j \in \mathbb{Z} .$$

Indeed, for $k - r \le j < k$, this follows from (10.2). For $j < k - r$, (10.3) merely says that $N_{j,T} = N_{j,U}$; similarly for $j \ge k$, it says $N_{j,T} = N_{j+1,U}$.

It follows from (10.3) that

$$S = \sum_{j \in \mathbb{Z}} c_{j,T} N_{j,T} = \sum_{j \in \mathbb{Z}} c_{j,T} \left(\alpha_{j+1} N_{j+1,U} + (1 - \alpha_j) N_{j,U}\right) .$$

Hence,

(10.4) $$c_{j,U} = \alpha_j c_{j-1,T} + (1 - \alpha_j) c_{j,T} , \quad j \in \mathbb{Z} .$$

In particular, $c_{j,U}$ is a convex combination of $c_{j-1,T}$ and $c_{j,T}$. From (10.4), it is easy to deduce the sign variation diminishing property of the B-spline representation.

Theorem 10.1 (Karlin [A-1968, p. 531]). *If $S \in \mathcal{S}_r(T)$ has the B-spline series $S = \sum_{j\in\mathbb{Z}} c_j N_j$ and the sequence (c_j) changes sign a finite number m of times, then S has at most m sign changes on $(-\infty, \infty)$.*

Proof. If U is a one knot refinement of T, as described above, then $(c_{j,U})$ changes sign no more often than $(c_{j,T})$. To see this, we construct a new sequence (d_j) from $(c_{j,T})$ by inserting $c_{j,U}$ between $c_{j-1,T}$ and $c_{j,T}$, $j \in \mathbb{Z}$. In view of (10.4), (d_j) changes sign no more often than $(c_{j,T})$. But $(c_{j,U})$ is a subsequence of (d_j) and therefore also changes sign no more often than $(c_{j,T})$, as desired. It follows by induction that for *any* refinement U of T obtained by inserting a finite number of knots, $(c_{j,U})$ changes sign no more often than $(c_{j,T})$.

Now let $x_1, \ldots, x_s$ be a finite strictly increasing sequence of real numbers. We want to show that $(S(x_i))_1^s$ has at most m sign changes. We can assume without loss of generality that x_i is not a knot, $x_i \neq t_j$ for all i, j. Let U be a refinement of T with the following property: each x_i appears in U *exactly* $r-1$ times. Then, there is exactly one B spline $N_{j,U}$ which is non-zero at x_i. The coefficient of this B-spline is $S(x_i)$ because from (3.4) $N_{j,U}(x_i) = 1$. Thus, $(S(x_i))_1^s$ is a subsequence of $(c_{j,U})$ and therefore $(S(x_i))_1^s$ changes sign no more often than $(c_{j,T})$. □

As an immediate corollary we obtain another proof of Theorem 8.4:

Corollary 10.2. *Each Schoenberg spline space $\mathcal{S}_r(T, I)$ is a weak Haar space.*

There is a useful geometric interpretation of (10.4). We define

$$t'_j := (t_{j+1} + \cdots + t_{j+r-1})/(r-1), \quad j \in \mathbb{Z} \tag{10.5}$$

to be the "knot averages" for T. If U is a one knot refinement of T, then the u'_j are related to the t'_j in the same way as (10.4):

$$u'_j = \alpha_j t'_{j-1} + (1-\alpha_j)t'_j, \quad j \in \mathbb{Z}. \tag{10.6}$$

We denote by $P := P(S,T)$ the continuous polygonal curve with vertices $(t'_j, c_{j,T})$, $j \in \mathbb{Z}$; P is called the "B-polygon" of S with respect to T. The relations (10.4) and (10.6) show that the vertices of the B-polygon $P(S,U)$ lie on the B-polygon $P(S,T)$. Hence $P(S,U)$ changes sign no more often than $P(S,T)$.

§11. Total Positivity of the B-spline Collocation Matrix

The entries in the collocation matrix C^* of (9.7) have a certain regularity. For example if the interpolation points X are distinct then the entries in C^* are all non-negative. Karlin [A-1968, Chapter 10] has established deeper properties of C^* using the concept of total positivity.

Let B be any $n \times m$ matrix, let $I : i_1 < \ldots, < i_k$, be a selection of row indices and let $J : j_1 < \cdots < j_k$ be a selection of column indices. Then the matrix $B(I,J) = (b_{ij})_{i\in I, j\in J}$ is a square submatrix of B. The determinant of $B(I,J)$ is called a *minor* of B. We say that B is *totally positive* if *all of its minors are non-negative (strictly totally positive if all of its minors are positive)*. In particular total positivity implies that the entries of B are non-negative. Any submatrix of a totally positive matrix is itself totally positive. A useful consequence of total positivity is that if B is a square matrix which is invertible then the entries in B^{-1} appear in a checkerboard pattern, that is, they alternate in sign along the rows and along the columns. This is an obvious consequence of Cramer's rule for computing the inverse of a matrix.

Let $T : t_1 \le \cdots \le t_{n+r}$ be a sequence of knots and $N_j := N_{j,T}$, $j = 1, \ldots, n$, be the corresponding B-splines. We take a sequence $X : x_1 < x_2 < \cdots < x_n$ of *distinct* interpolation points and have the collocation matrix of Lagrange interpolation

$$C^* := (N_j(x_i))_{i,j=1}^n. \tag{11.1}$$

To prove that C^* is totally positive, we shall follow de Boor and DeVore [1985] and use knot refinements as introduced in the last section. We select the index sets of length k, $I : i_1 < \cdots < i_k$ and $J : j_1 < \cdots < j_k$, and let

$$B := (N_j(x_i))_{i\in I, j\in J}. \tag{11.2}$$

We shall show that $\det(B) \ge 0$. For this purpose, we consider an arbitrary knot refinement U obtained from T by adjoining a finite number of knots and show that $\det(B)$ is a non-negative linear combination of minors $\det(B_H)$, with

$$B_H := (N_{j,U}(x_i))_{i\in I, j\in H},$$

where $1 \le h_1 \le \cdots \le h_k \le n$ is an arbitrary sequence of integers of length k. We claim that

$$\det(B) = \Sigma^+ \gamma_H \det(B_H), \quad \gamma_H \ge 0, \tag{11.3}$$

where the sum Σ^+ extends over all *strictly increasing* sequences H.

To prove (11.3), we consider first the special case when U is a one knot refinement of T as in §10; we insert the new knot between t_{k-1} and t_k. Using (10.3), we can replace the values $N_{j,T}$ in all columns of the matrix (11.2) by $(1-\alpha_j)N_{j,U} + \alpha_{j+1}N_{j+1,U}$ with $\alpha_{j+1} \ge 0$ and $1-\alpha_j \ge 0$. Since a determinant is a linear function of its columns, this yields

(11.4) $$\det(B) = \sum \gamma_H \det(B_H)\,, \quad \gamma_H \geq 0\,,$$

the sum taken over all sequences H of length k that are non-decreasing. If two entries in H are equal then the corresponding columns in B_H are identical and hence $\det(B_H) = 0$. Therefore, (11.4) reduces to (11.3). The general case (11.3) now follows by inserting one knot at a time to go from T to U.

Theorem 11.1. *If the interpolation points X are distinct, then the collocation matrix* (11.1) *is totally positive.*

Proof. We need to show that each matrix B of (11.2) has $\det(B) \geq 0$. Let $U := (u_i)_{-\infty}^{\infty}$ be a knot refinement of T with the property that U has at least r knots between x_i and x_{i+1}, for each $1 \leq i < k$. Then for $i = 1, \ldots, k$, we have

(11.5) $$\text{if } N_{j,U}(x_i) \neq 0, \text{ then } N_{j,U}(x_{i'}) = 0\,, \quad i \neq i'\,.$$

Now let B_H be any submatrix appearing in (11.3). Because of (11.5), B_H has at most one non-zero entry in each column and this entry is positive. Since the row indices of these non-zero entries do not decrease with increasing column index, we have $\det(B_H) \geq 0$. Hence (11.3) gives, $\det(B) \geq 0$. □

If the interpolation points are not distinct, then the collocation matrix C^* of (9.7) is not totally positive because derivatives of B-splines can be negative.

§ 12. Problems

12.1. For $0 \leq k \leq n$, $N(x; 0, \ldots, 0, 1, \ldots, 1) = \binom{n}{k} x^k (1-x)^{n-k}$, $0 < x < 1$ if 0 appears $n - k + 1$ times and 1 appears $k + 1$ times in the definition of N.

12.2. For $r \geq 3$, construct an increasing sequence $(t_j')_{-\infty}^{\infty}$ such that the t_j' are not the knot averages (10.4) of any knot sequence T.

12.3. If $t_j := \cos(r-j)\pi/r$, $j = 0, \ldots, r$, then the B-spline $M(x) := M(x; t_0, \ldots, t_r)$ is a perfect spline, that is $\left|M^{(r-1)}(x)\right| = \text{const.}$, $-1 < x < 1$, $x \neq t_j$, $j = 1, \ldots, r-1$.

12.4. If $S \in \mathcal{S}_r(T, I)$ is continuous, $I = [a, b]$, then for the knot refinements U of T, the B-polygons $P(S, U)$ converge uniformly to S on I as $\delta_U \to 0$ where $\delta_U := \max_j |u_{j+1} - u_j|$ is the mesh length of $U := (u_j)$.

12.5. With the assumptions of Problem 12.4, there is a constant $C > 0$ such that $\|S - P(S, U)\| \leq C\|S''\|\delta_U^2$ in the uniform norm on I whenever S is twice continuously differentiable.

12.6. If $x_0 < x_1 < \cdots < x_r$, then the B-spline $M(x) := M(x, x_0, \ldots, x_r)$ is bell-shaped, that is M' has exactly one sign change. More generally $M^{(k)}$ changes sign k times, $k = 1, \ldots, r-1$.

12.7. In Theorem 6.1, the points X can be repeated up to m times if such repetition corresponds to prescribing derivatives as in (6.3).

12.8. For $I = [a, b]$, $E(f, \mathcal{S}_{r+1}(T, I))_\infty \leq C\delta_T E(f', \mathcal{S}_r(T, I))_\infty$, whenever $f' \in C(I)$.

12.9. Use the recurrence formula for the B-splines (2.5) to derive a formula for the coefficients $\alpha_{i,j}^{(k)}$ in the Oslo algorithm of the Note 13.7 (Prautzsch [1984]).

§ 13. Notes

13.1. Different possible representations of the dual functions c_j for the B-spline basis (N_j) have been discussed by Lyche and Schumaker (see Schumaker [A-1981, Chapter 4]). We have already seen that the point ξ_j in the representation (3.6) can be chosen arbitrarily in the interval (t_j, t_{j+r}). More generally, if $d\mu_j$ are any positive measures of unit mass supported on (t_j, t_{j+r}), then for $j \in \Lambda$,

$$c_j(f) := \int_{-\infty}^{\infty} \left[\sum_{\nu=0}^{r-1} (-1)^\nu g_j^{(r-\nu-1)}(t) f^{(\nu)}(t)\right] d\mu_j(t)\,,$$

are also dual functionals to the $N_{j,r}$.

There are also representations which do not involve derivatives of f. We let $G_j(t) := \phi_j(t) g_j(t)$ where ϕ_j are $r-1$ times continuously differentiable functions which vanishes for $t < t_j$ and is equal to one for $t > t_{j+r}$. Then,

$$c_j(f) := \int_{-\infty}^{\infty} f(t) G_j^{(r)}(t) dt\,, \quad j \in \Lambda\,,$$

is another representation for the dual functionals.

13.2. Perfect splines occur often in extremal problems. Karlin [1973] has shown that the extremal problem $(*)$ of §6 with the L_2 norm replaced by the L_∞ norm has as one of its solutions (uniqueness of solutions does not hold for $p = \infty$) a perfect spline of degree m with fewer than $n - m$ knots (see Lorentz, Jetter, and Riemenschneider [A-1983, p. 217].

13.3. The original proof by Kolmogorov of (7.1) was different from ours. It was based on "Kolmogorov's comparison theorem" below. It is of independent interest and importance (see Korneichuk [A-1976] for a proof and applications).

Theorem. *If a function $F \in W_\infty^m(\mathbb{R})$ of period 2 satisfies* (7.4) *and in addition at some points x_0, x_1 one has $F(x_0) = \mathcal{E}_m(x_1)$, then $\left|F'(x_0)\right| \leq \mathcal{E}_m'(x_1)$.*

13.4. For the number of zeros of an individual spline S there is a stronger inequality than (8.2). We consider maximal vertical sequences (t_i^*, j), $k_1 \leq j \leq k_2$ of singular points in Φ, that have an odd number of elements and are supported in the sense that there are singular points $(t_{i_1}^*, j_1)$, $(t_{i_2}^*, j_2)$ in Φ for which $t_{i_1}^* < t_i^* < t_{i_2}^*$, $j_1, j_2 < k$. If $\gamma(S)$ is the number of such sequences, then

$$Z(S) \leq \sigma - L(\Phi) - r - 1 - \gamma(S)$$

(see Jetter, Lorentz, and Riemenschneider [A-1983]; weaker forms of this theorem go back to Birkhoff [1906]).

13.5. There exist generalizations of Theorem 9.2 which assert the solvability of *Birkhoff interpolation problems* for splines in a Schoenberg space, and in more general spaces, for instance by Goodman (see the book of Lorentz, Jetter, and Riemenschneider [A-1983, Chapter 14]).

13.6. The B-polygons, defined in §10 after (10.5)are sometimes used in computer aided geometric design as a means of generating smooth curves from a given ordered set of planar points (called control points) (x_i, y_i), $i = 0, \ldots, n$. The curve is not required to pass through these points but should reflect their geometric properties. A typical way of generating such curves is to begin with a polygon $P = P_0$ (called the control polygon) which has these points as its vertices and then to produce a new polygon P_1 which has a new and large number of vertices generated in some algorithmic (usually geometric) way from the vertices of P_0. This process is then repeated to generate a sequence of polygons $P_0, P_1, \ldots$. The procedure stops when P_n appears as a smooth curve when it is presented on a computer monitor.

In some algorithms, the original polygon P_0 can be thought of as a B-polygon $P(S,T)$ of a spline S for the knot sequence T. Then $P_1 = P(S,U)$ is the B polygon of S for some refined knot sequence U. This is the case, for example, in the Chaikan [1974] algorithm which can be viewed as coming from knot insertion for quadratic splines.

13.7. If U is a knot refinement of T, then any spline $S \in \mathcal{S}_r(T, \mathbb{R})$ can be represented in two B-spline series (10.1). In particular,

$$N_{i,T} = \sum_j \alpha_{i,j}^{(r)} N_{j,U} .$$

It follows that

$$c_{j,U} = \sum_i \alpha_{i,j}^{(r)} c_{i,T} .$$

The coefficients $\alpha_{i,j}^{(r)}$ are sometimes called *discrete* B-splines. Cohen, Lyche, and Riesenfeld [1984] have given a recurrence formula (known as the Oslo Algorithm) which can be used to compute these coefficients. One has

$$\alpha_{i,j}^{(1)} := \begin{cases} 1 & \text{if } t_i \le u_j < t_{i+1} \\ 0, & \text{otherwise.} \end{cases}$$

For $r > 1$, the coefficients $\alpha_{i,j}^{(r)}$ can be computed from the recursive formulas:

$$\alpha_{i,j}^{(k)} = \frac{u_{j+k-1} - t_i}{t_{i+k-1} - t_i} \alpha_{ij}^{(k-1)} + \frac{t_{i+k} - u_{j+k-1}}{t_{i+k} - t_{i+1}} \alpha_{i+1,j}^{(k-1)} .$$

13.8. There is a simple algorithm given by de Boor (see his book [A-1978, Chapter 10] for computing the values of $S(x) = \Sigma c_j N_j(x)$. For $x \in [t_i, t_{i+1}]$, it begins with the r coefficients $c_j^{(1)} := c_j$, $j = i - r + 1, \ldots, i$ for which $N_j(x) \ne 0$ and then generates recursively

$$c_j^{(k+1)} := \left(1 - \alpha_j^{(k)}\right) c_{j-1}^{(k)} + \alpha_j^{(k)} c_j^{(k)} ; \quad \alpha_j^{(k)} := \frac{x - t_j}{t_{j+r-k} - t_j} .$$

This gives a triangular array whose last entry is $c_i^{(r)} = S(x)$.

13.9. To generate a graph $y = S(x)$ of a spline $S \in \mathcal{S}_r(T, I)$, $I = [a,b]$, it is not necessary to compute values of S. Instead, one can use the B-polygons $P(S,U)$. The graph of such a B-polygon will approximate well the graph of S if U is a fine enough refinement of T. To determine $P(S,U)$, one can either use the formulas (10.3), (10.5) and insert one knot at a time to go from T to U or one can use the Oslo algorithm of Note **13.7.**

13.10. Theorem 11.1 is not true in the case of multiple interpolation points. However, if M is a minor of A formed from rows $i_1, \ldots, i_m$ and columns $j_1, \ldots, j_m$ with the property

$$i_{\nu-1} < i_\nu - 1 \text{ implies } x_{i_\nu - 1} < x_{i_\nu} , \quad \nu = 1, \ldots, m ,$$

then $M \ge 0$. In addition, M is positive if and only if $x_{i_\nu} \in \operatorname{supp}(N_{j_\nu})$, $\nu = 1, \ldots, m$. This can be derived from (11.2) by examining exactly which coefficients γ_H are positive (see de Boor and DeVore[1985]). In this way, we obtain another proof of the Karlin-Ziegler theorem.

Chapter 6. K-Functionals and Interpolation Spaces

§ 1. K-Functionals

Moduli of smoothness $\omega_r(f,t)$ and K functionals $K(f,t)$ are both functions of the real parameter $t \geq 0$ which express some intrinsic properties of the function f. The former give direct information about the smoothness of f. The latter have been introduced by Lions and Peetre and in their present form (1.1) by Peetre [B-1963] as a basis for his theory of *interpolation of operators.* Actually, the K-functionals express some approximation properties of f (see below). Thus, they turn out to play an equally important role in approximation. For the application of K-functionals to interpolation theorems of operators compare Chapter 5 of the book Bennett and Sharpley [B-1988]. The book of Butzer and Berens [A-1967] made clear their importance in approximation. This has been later amplified by the fact that in many situations (see §2 and also §6) $K(f,t)$ and $\omega_r(f,t)$ are equivalent.

We begin by defining the K-functional in their simplest setting and later extend this definition in several directions. Let X_i, $i = 0,1$, be two Banach spaces, with X_1 continuously embedded in $X_0 : X_1 \subset X_0$. We define the K-functional for $f \in X_0$ by

$$K(f,t) := K(f,t;X_0,X_1) := \inf_{g \in X_1} \{\|f-g\|_{X_0} + t\|g\|_{X_1}\}, \quad t \geq 0. \tag{1.1}$$

This quantity expresses some approximation properties of f. Indeed, the inequality $K(f,t) < \varepsilon$ for some $t > 0$ implies that f can be approximated with error $\|f-g\| < \varepsilon$ in X_0 by an element $g \in X_1$, whose norm is not too large: $\|g\|_{X_1} < \varepsilon t^{-1}$.

Often one needs K-functionals in more general situations, with the following deviations from the simplest case just described: (1) the X_i, $i = 0,1$, are only quasi-normed spaces; (2) X_0 is not embedded in X_1; and (3) instead of the quasi-norm $\|g\|_{X_1}$, one uses in (1.1) a quasi-semi norm $|g|_{X_1}$.

In quasi-normed spaces, defined in §1 of Chapter 2, instead of the triangle inequality, we have

$$\|f+g\|_X \leq C(\|f\|_X + \|g\|_X). \tag{1.2}$$

We have seen in §1 of Chapter 2, that there is then an equivalent quasi-norm on X such that for some $\mu > 0$ (and then for all smaller positive μ)

$$\|f+g\|_X^\mu \leq \|f\|_X^\mu + \|g\|_X^\mu. \tag{1.3}$$

We shall always assume that (1.3) is satisfied when working with quasi-normed spaces.

In case (2), one postulates the existence in X_0, X_1 of a common convergence, weaker than both quasi-norm convergences. As an example, for the spaces L_1, L_∞ on $\mathbb{R}$, a weaker convergence is convergence in measure. Two quasi-normed, complete, linear spaces X_0, X_1 are called a *pair,* (X_0, X_1), if each of them is continuously embedded in a linear Hausdorff topological space $\mathcal{X}$. In the example of L_1, L_∞, $\mathcal{X}$ is the space M of all measurable a.e. finite functions on $\mathbb{R}$. For such a pair, we can define new spaces $X_0 + X_1$, $X_0 \cap X_1$. The first space consists of elements $f \in X$ which possess a representation $f = f_0 + f_1$, $f_i \in X_i$, $i = 0,1$. We define

$$\|f\|_{X_0+X_1} := \inf_{f=f_0+f_1} \{\|f_0\|_{X_0} + \|f_1\|_{X_1}\}. \tag{1.4}$$

Similarly, $X_0 \cap X_1$ is a complete linear space with the quasi-norm

$$\|f\|_{X_0 \cap X_1} = \max(\|f\|_{X_0}, \|f\|_{X_1}). \tag{1.5}$$

We leave it to the reader to show that these are quasi-norms (or norms if X_0, X_1 are normed) and that the spaces $X_0 + X_1$ and $X_0 \cap X_1$ are complete with respect to their topologies (see also Bennett and Sharpley [B-1988, p. 98]).

These spaces satisfy the continuous embeddings:

$$X_0 \cap X_1 \subset X_0, X_1 \subset X_0 + X_1, \tag{1.6}$$

which follow from the simple inequalities:

$$\|f\|_{X_0+X_1} \leq \|f\|_{X_0}, \|f\|_{X_1} \leq \|f\|_{X_0 \cap X_1}. \tag{1.7}$$

We can now extend the definition (1.1) to the case of an arbitrary pair of quasi-normed spaces. If $f \in X_0 + X_1$ and $t \geq 0$, we put

$$K(f,t) := K(f,t;X_0,X_1) := \inf_{f=f_0+f_1} \{\|f_0\|_{X_0} + t\|f_1\|_{X_1}\}. \tag{1.8}$$

For a fixed $t > 0$, $K(\cdot,t)$ is equivalent to $\|\cdot\|_{X_0+X_1}$; indeed

$$\min(1,t)\|f\|_{X_0+X_1} \leq K(f,t) \leq \max(1,t)\|f\|_{X_0+X_1}. \tag{1.9}$$

There is a further generalization in which $\|f_1\|_{X_1}$ in (1.8) is replaced by a quasi-seminorm $|f|_{X_1}$ on X_1 with some null space $\mathcal{N} \subset X_1$:

$$K(f,t;X_0,X_1) := \inf_{f=f_0+f_1} \{\|f_0\|_{X_0} + t|f_1|_{X_1}\}. \tag{1.10}$$

For fixed $t > 0$, $K(\cdot,t)$ of (1.10) is a quasi-seminorm which is equivalent to $\|f\|_{X_0+X_1}$.

In particular, we shall need this generalization for the spaces $X_0 = L_p$, $X_1 = W_p^r$ on A (where A is one of the standard intervals or $\mathbb{T}$) and the semi-norm $|g|_{W_p^r} := \|g^{(r)}\|_p$. In this case,

$$K\left(f,t;L_p,W_p^r\right) := \inf_{g\in W_p^r}\left\{\|f-g\|_p + t\|g^{(r)}\|_p\right\} . \tag{1.11}$$

If $A=[a,b]$, the null space $\mathcal{N}$ is the subspace of all polynomials of degree $< r$. In particular, $\mathcal{N}=\{0\}$ if $|A|=\infty$, $1\le p<\infty$ and $\mathcal{N}$ consists of constants if $|A|=\infty$, $p=\infty$ or if $A=\mathbb{T}$.

Proposition 1.1. *The K-functional* (1.8) *or* (1.10) *has the properties:*

(i) *As a function of $t\ge 0$, $K(f,t)$ is increasing, concave and continuous. It is also subadditive: $K(f,t_1+t_2)\le K(f,t_1)+K(f,t_2)$.*

(ii) *If X_0, X_1 are Banach spaces, then as a function of f for each fixed $t>0$, $K(f,t)$ is a seminorm on X_0+X_1.*

(iii) *If X_0, X_1 are quasi-normed spaces, then as a function of f, for each fixed $t>0$, $K(f,t)$ is a quasi-seminorm; for any $f,g\in X_0+X_1$:*

$$K(f+g,t)\le C\left(K(f,t)+K(g,t)\right) . \tag{1.12}$$

Proof. It is clear that $K(f,t)$ is continuous and increasing; it is concave by statement (6.9) of Chapter 2. Subadditivity follows from concavity (see Example 1 of §6 of Chapter 2). We leave the simple verification of (ii) and (iii) to the reader. □

The K-functional is not symmetric with respect to X_0, X_1; instead, from the definition, we have

$$K\left(f,t;X_1,X_0\right) = t\,K\left(f,1/t;X_0,X_1\right) .$$

In the case that $X_0=X_1=X$ is a normed space, we have

$$K(f,t;X,X)=\min(1,t)\|f\|_X\,,\quad t\ge 0 .$$

In a Banach function space X (see §1 of Chapter 2), $\|\phi f\|=\|f\|$ if $|\phi(x)|=1$, a.e. Hence if X_0, X_1 are of this type, then

$$\inf_{g\in X_1}\{\|\phi f-g\|_{X_0}+\|g\|_{X_1}\} = \inf_{g_1\in X_1}\{\|f-g_1\|_{X_0}+\|g_1\|_{X_1}\}$$

and therefore, $K(\phi f,t)=K(f,t)$. In particular,

$$K(|f|,t)=K(f,t)\,,\quad t\ge 0 . \tag{1.13}$$

The original application of K-functionals was in describing the mapping properties of linear operators U. If U maps the quasi-normed spaces X_i into Y_i with norm M_i, $i=0,1$, then for $f\in X_0$ and any $g\in X_1$,

$$\begin{aligned} K\left(Uf,t;Y_0,Y_1\right) &\le \|U(f-g)\|_{Y_0}+t\|Ug\|_{Y_1}\\ &\le M_0\|f-g\|_{X_0}+M_1t\|g\|_{X_1} . \end{aligned}$$

With $M:=\mathrm{Max}(M_0,M_1)$, it follows that

$$K\left(Uf,t;Y_0,Y_1\right)\le M_0K\left(f,M_1t/M_0;X_0,X_1\right)\le M\,K\left(f,t;X_0,X_1\right) . \tag{1.14}$$

The same estimate applies if $|Uf|_{Y_1}\le M_1|f|_{X_1}$ and the K-functionals are defined by (1.10). Several applications of (1.14) are given later in this chapter.

As an example, we compute the K-functional for the pair $(L_p(A),L_\infty(A))$, both spaces equipped with the Lebesgue measure μ, in terms of the decreasing rearrangement f^* ((2.2) of Chapter 2).

Theorem 1.2. *If $0<p<\infty$, then for each $f\in L_p+L_\infty$, we have*

$$K(f,t,L_p,L_\infty)\sim\left(\int_0^{t^p} f^*(s)^p ds\right)^{1/p}\,,\quad t\ge 0 . \tag{1.15}$$

Equality holds in (1.15) *if $p=1$.*

Proof. We fix p and $t>0$ and define $\Lambda(f):=\left(\int_0^{t^p}f^*(s)^pds\right)^{1/p}$. Let $f=f_0+f_1$, with $f_0\in L_p$, $f_1\in L_\infty$. If $p=1$, Λ is subadditive and therefore

$$\Lambda(f)\le\Lambda(f_0)+\Lambda(f_1)\le\|f_0^*\|_1+t\|f_1^*\|_\infty=\|f_0\|_1+t\|f_1\|_\infty .$$

Taking an infimum over all such decompositions, we obtain that the right side of (1.15) does not exceed the left. For $p\ne 1$, in place of the subadditivity of Λ, we use the weak subadditivity of f^* (see 2(vii) of Chapter 2):

$$f^*(s)\le f_0^*(s/2)+f_1^*(s/2)\,,\quad s>0 .$$

This gives

$$\Lambda(f)\le C\left(\Lambda(f_0(\cdot/2))+\Lambda(f_1(\cdot/2))\right)\le C\left\{2^{1/p}\|f_0\|_p+t\|f_1\|_\infty\right\} ;$$

in particular $C=1$ if $p\ge 1$. We obtain that the right side of (1.15) does not exceed a multiple of the left.

For the inequality in the other direction, we can, because of (1.13), assume that $f\ge 0$. For a fixed $t>0$, we define $M:=f^*(t^p)$, $g(x):=\min(f(x),M)$ and $E:=\{x: f(x)>M\}$. Then, $\left((f-M)\chi_E\right)^*(s)=f^*(s)-M$, $0\le s\le\mu(E)$. If $\Lambda(f)<\infty$, then $\mu(E)\le t^p$ and $f-g\in L_p$. Therefore,

$$\begin{aligned} \|f-g\|_p^p+t^p\|g\|_\infty^p &\le \int_E (f-M)^p d\mu + t^pM^p\\ &\le \int_0^{\mu(E)}\left(f^*(s)-M\right)^p ds+t^pM^p . \end{aligned} \tag{1.16}$$

When $p=1$, the right side of (1.16) equals $\int_0^t f^*(s)ds$ because $f^*(s)=M$, $\mu(E)<s\le t$. In this case, the left side of (1.15) does not exceed the right. When $p\ne 1$, both terms on the right in (1.16) do not exceed $\int_0^{t^p}f^*(s)ds$ and we obtain that the left side of (1.15) does not exceed a multiple of the right. □

If U is a bounded linear operator from L_1, L_∞ into themselves then U satisfies (1.14) for these spaces. From (1.15), we obtain $U(f) \prec CMf$, where $\prec$ is the Hardy-Littlewood relation (2.8) of Chapter 2. If X is any rearrangement-invariant space on A, we get from Theorem 2.2 of Chapter 2 that $\|Uf\|_X \le CM\|f\|_X$, $f \in X$. This yields a new proof of Theorem 4.4 of Chapter 2.

We shall need still another extension of the definition of the K-functional. If X is a complete quasi-normed space with quasi-norm $\|f\|$, then it also has these properties with the quasi-norm on X replaced by a power quasi-norm $\|f\|^q$ with $0 < q < \infty$. We denote this space by $Y := X^q$. This is not a quasi-norm since homogeneity fails when $q \ne 1$. It is of some interest to investigate the effect of this operation on the K-functional of a pair (X_0, X_1). Let $Y_i := X_i^{q_i}$, $i = 0, 1$. We define the K-functional for this pair as in (1.10) with the quasi-norms now replaced by the respective power quasi-norms. Then (Y_0, Y_1) can be embedded in a Hausdorff space $\mathcal{X}$ and the linear spaces $Y_0 + Y_1$ and $X_0 + X_1$ are identical. We compare

$$K(f,t) := K(f,t;X_0,X_1)\,, \quad \overline{K}(f,s) := K(f,s;Y_0,Y_1)\,. \tag{1.17}$$

Since K is increasing and concave, $t^{-1}K(f,t)$ is decreasing. Hence $t^{q_1}K(f,t)^{q_0-q_1}$ is, for $t \ge 0$, continuous, strictly increasing, vanishes for $t = 0$ and by (1.9), tends to infinity with t. It follows that each $s \ge 0$ has a unique representation

$$s = t^{q_1}K(f,t)^{q_0-q_1}\,, \quad t \ge 0\,. \tag{1.18}$$

Proposition 1.3. *For each $f \in X_0 + X_1$, $t \ge 0$, and s given by (1.18), the functions $\overline{K}(f,s)$ and $K(f,t)^{q_0}$ are equivalent; with $q := \max(q_0, q_1)$*

$$2^{-q}K(f,t)^{q_0} \le \overline{K}(f,s) \le 2K(f,t)^{q_0}\,. \tag{1.19}$$

Proof. For fixed f and t, we put $K := K(f,t)$, $\overline{K} := \overline{K}(f,s)$. If $\varepsilon > 0$, let $f = f_0 + f_1$ with $\|f_0\|_{X_0} + t\|f_1\|_{X_1} \le K + \varepsilon$. Then

$$\overline{K} \le \|f_0\|_{X_0}^{q_0} + s\|f_1\|_{X_1}^{q_1} \le (K+\varepsilon)^{q_0} + t^{q_1}K^{q_0-q_1}t^{-q_1}(K+\varepsilon)^{q_1}\,.$$

If we let $\varepsilon \to 0$, we obtain the right inequality in (1.19). Conversely, let $f = f_0 + f_1$ with $\|f_0\|_{X_0}^{q_0} + s\|f_1\|_{X_1}^{q_1} \le \overline{K} + \varepsilon$. Then, $\|f_1\|_{X_1} \le s^{-1/q_1}(\overline{K}+\varepsilon)^{1/q_1}$, hence from (1.18)

$$K \le \|f_0\|_{X_0} + t\|f_1\|_{X_1} \le (\overline{K}+\varepsilon)^{1/q_0} + K^{1-q_0/q_1}(\overline{K}+\varepsilon)^{1/q_1}\,. \tag{1.20}$$

If we let $\varepsilon \to 0$ then we obtain the left inequality in (1.19) because K is less than either twice the first term or twice the second term on the right side of (1.20). □

§ 2. K-Functionals and Moduli of Smoothness

There is a close connection between K functionals and the functions ω, ω_r. We begin with the modulus of continuity ω. The first theorem is in terms of the concave majorant $\overline{\omega}$ of ω (see §6 of Chapter 2); it satisfies $\omega(f,t) \le \overline{\omega}(f,t) \le 2\omega(f,t)$.

Theorem 2.1 (Korneichuk [1961]. *For each function $f \in C(A)$, $A = [a,b]$ or $A = \mathbb{T}$, and the concave majorant $\overline{\omega}$ of its modulus of continuity ω,*

$$K(f,t;C,\mathrm{Lip}\,1) = \tfrac{1}{2}\overline{\omega}(f,2t)\,, \quad t \ge 0\,. \tag{2.1}$$

Proof (by Mitjagin and Semenov [1977]). For arbitrary $f \in C(A)$, $g \in \mathrm{Lip}\,1$, and $t \ge 0$, $\omega(f,2t) \le \omega(f-g,2t) + \omega(g,2t) \le 2\|f-g\|_C + 2t|g|_{\mathrm{Lip}\,1}$; hence $\omega(f,2t) \le 2K(f,t)$, and since K is concave,

$$\tfrac{1}{2}\overline{\omega}(f,2t) \le K(f,t)\,. \tag{2.2}$$

To prove the reverse inequality, let $f \in C(A)$ and let $t > 0$ be fixed. For the function $\overline{\omega}$ and the point $2t$, let $l(s) := \overline{\omega}(f,2t) + M(s-2t)$, where M is a constant, be a supporting linear function of $\overline{\omega}$ at $s = 2t$. Since $\overline{\omega}(f,s) \le l(s)$, $s \ge 0$,

$$\delta := \tfrac{1}{2}\sup_{s>0}\left[\overline{\omega}(f,s) - Ms\right] = \tfrac{1}{2}\left[\overline{\omega}(f,2t) - 2Mt\right]\,. \tag{2.3}$$

To find a $g \in \mathrm{Lip}\,1$ corresponding to f, we put, for each $y \in A$, $f_y(x) := f(y) - M|x-y| - \delta$, $x \in A$. Obviously $f_y \in \mathrm{Lip}\,1$ and $|f_y|_{\mathrm{Lip}\,1} \le M$. We define $g(x) := \sup_{y\in A} f_y(x)$. From the inequality $f_y(x_1) \le f_y(x_2) + M|x_1 - x_2|$, we derive a similar inequality for g, hence $|g|_{\mathrm{Lip}\,1} \le M$. For this function g,

$$g(x) - f(x) = \sup_y\,[f(y) - f(x) - M|x-y| - \delta] \ge -\delta\,. \tag{2.4}$$

On the other hand, by (2.3) for $x, y \in A$, $f(y) - f(x) - M|x-y| \le \omega(f,|x-y|) - M|x-y| \le 2\delta$, so that

$$g(x) - f(x) \le \delta\,. \tag{2.5}$$

Hence, $\|f-g\|_C \le \delta$, and we obtain

$$K(f,t) \le \|f-g\|_C + t|g|_{\mathrm{Lip}\,1} \le \delta + Mt = \tfrac{1}{2}\overline{\omega}(f,2t)\,.$$ □

Corollary 2.2. *Let $A = [a,b]$ or $\mathbb{T}$ and let U be a linear operator which is bounded on $C(A)$ and $\mathrm{Lip}\,1$ with norms not exceeding M on both spaces. Then for $f \in C(A)$*

$$\overline{\omega}(Uf,t) \le M\overline{\omega}(f,t) \quad t > 0\,. \tag{2.6}$$

In particular, U boundedly maps $\mathrm{Lip}\,\alpha$ into itself for each $0 < \alpha \le 1$.

Proof. The inequality (2.6) is (1.14). Since $\omega(f,t) \le \overline{\omega}(f,t) \le 2\omega(f,t)$, (see Lemma 6.1 of Chapter 2), (2.6) implies that T maps $\operatorname{Lip}\alpha$ to $\operatorname{Lip}\alpha$, $0<\alpha\le 1$. $\square$

For the modulus of smoothness, we can only show that $\omega_r(f,t)_p$ and the functional $K\left(f,t;L_p,W_p^r\right)$ are equivalent up to constants (Johnen [1972], Peetre [B-1963]; we follow the approach of Johnen-Scherer [1976]). Let $I_1 = [a_1,b_1]$, $I_2=[a_2,b_2]$, $a_1<a_2<b_1<b_2$ and let $I=I_1\cup I_2$ and $J=I_1\cap I_2$. For the K functional in question we have

Lemma 2.3. *For some constant C, depending only on r and the ratio $|I|/|J|$,*

$$K(f,t)(I) \le C\{K(f,t)(I_1)+K(f,t)(I_2)\}\,, \quad t\le |I|\,. \tag{2.7}$$

Proof. Let ϕ be the function of Lemma 5.1 of Chapter 2 for $J=[c,d]$ and $\delta := |J|/4$. For an appropriate constant $\gamma \le C/|J|$, the function $\psi(x) := \gamma\int_{-\infty}^x \phi(t)dt$ vanishes for $x\le c$, is increasing on J and is identically 1 for $x\ge d$. Also, $\left|\psi^{(k)}(x)\right| \le C|J|^{-k}$, $k=0,1,\dots,r$. If $f\in W_p^r(I)$, $t\le |I|$, let $g_i\in W_p^r(I_i)$, $i=1,2$, be chosen arbitrarily; we put $g:=(1-\psi)g_1+\psi g_2$. Then, in the L_p-norm,

$$\|f-g\|(I) \le \|f-g_1\|(I_1)+\|f-g_2\|(I_2)\,. \tag{2.8}$$

We need similar estimates for $t^r\|g^{(r)}\|(I)$. On $I\backslash J$, the function ψ is either 0 or 1 and $g^{(r)}=(1-\psi)g_1^{(r)}+\psi g_2^{(r)}$. Hence,

$$t^r\|g^{(r)}\|(I\backslash J) \le t^r\left\{\|g_1^{(r)}\|(I_1)+\|g_2^{(r)}\|(I_2)\right\}\,. \tag{2.9}$$

On J, we apply to $g=g_1+\psi(g_2-g_1)$ the Leibniz rule and find

$$\begin{aligned} t^r\|g^{(r)}\|(J) &\le Ct^r\left\{\|g_1^{(r)}\|(J)+\max_{0\le k\le r}\left[|J|^{-(r-k)}\|g_2^{(k)}-g_1^{(k)}\|(J)\right]\right\} \\ &\le C\left\{t^r\|g_1^{(r)}\|(I_1)+\max_{0\le k\le r}\left(t^k\|g_2^{(k)}-g_1^{(k)}\|(J)\right)\right\}, \end{aligned} \tag{2.10}$$

provided $t\le |I|$. By Theorem 5.6 of Chapter 2,

$$\begin{aligned} t^k\left\|g_2^{(k)}-g_1^{(k)}\right\|(J) &\le C\left\{\|g_2-g_1\|(J)+t^r\left\|g_2^{(r)}-g_1^{(r)}\right\|(J)\right\} \\ &\le C\left\{\|f-g_1\|(I_1)+t^r\left\|g_1^{(r)}\right\|(I_1)+\|f-g_2\|(I_2)+t^r\left\|g_2^{(r)}\right\|(I_2)\right\}. \end{aligned}$$

Substituting this into (2.10), and using also (2.8) and (2.9), we obtain that $\|f-g\|(I)+t^r\|g^{(r)}\|(I)$ does not exceed a multiple of the right hand expression in the last inequality. By the definition of the K functional this establishes (2.7). $\square$

Theorem 2.4 (Johnen [1972]). *Let $A=\mathbb{R}$, $\mathbb{R}_+$, $\mathbb{T}$ or $[a,b]$. For $1\le p\le\infty$ and $r=1,2,\dots$ there exist constants C_1, $C_2>0$ which depend only on r such that for all $f\in L_p$*

$$C_1\omega_r(f,t)_p \le K\left(f,t^r;L_p,W_p^r\right) \le C_2\omega_r(f,t)_p\,, \quad t>0\,. \tag{2.11}$$

Proof. For an arbitrary $g\in W_p^r$, from the elementary properties of the modulus of smoothness (7.3), (7.5) and (7.12) of Chapter 2, we obtain

$$\omega_r(f,t)_p \le \omega_r(f-g,t)_p+\omega_r(g,t)_p \le 2^r\|f-g\|_p+t^r\|g^{(r)}\|_p\,.$$

This shows that the left inequality in (2.11) is valid with $C_1=2^{-r}$.

To obtain the right inequality, we compare f with the function

$$g(x) := f(x)+(-1)^{r+1}\int_{-\infty}^{\infty}\Delta_{tu}^r(f,x)M(u)du \tag{2.12}$$

which depends on $t>0$. Here $M(u)=M(0,\dots,r;u)$ is the B-spline of Chapter 5 of order r with knots $0,1,\dots,r$. The function g is defined on A if A is $\mathbb{R}$, $\mathbb{R}_+$ or $\mathbb{T}$. If $A=[a,b]$, it is certainly defined if $x\in I_1 := \left[a,b-\frac{b-a}{4}\right]$ and if $t\le t_0 := (b-a)/4r^2$. Excluding the case $A=[a,b]$ for the moment, we have

$$\|f-g\|_p(A) \le \int_{-\infty}^{\infty}\|\Delta_{tu}^r(f,\cdot)\|_p(A)\,M(u)du \le \omega_r(f,rt)_p \le r^r\omega_r(f,t)_p\,, \tag{2.13}$$

since M has support $[0,r]$ and integral one.

To estimate $g^{(r)}$, let F be one of the r-th integrals of f on A. The right side of (2.12) is a linear combination of the integrals

$$\begin{aligned} \int_{-\infty}^{\infty} f(x+jtu)M(u)du &= \int_{-\infty}^{\infty} f(x+u)M\left((jt)^{-1}u\right)(jt)^{-1}du \\ &= (jt)^{-r}\Delta_{jt}^r(F,x)\,, \quad j=1,\dots,r\,, \end{aligned}$$

by (2.2) of Chapter 5 (because $M\left((jt)^{-1}u\right)(jt)^{-1}=M(u;0,jt,\dots,rjt)$). The r-th derivative of this is $(jt)^{-r}\Delta_{jt}^r(f,x)$. We obtain in this way

$$g^{(r)}(x)=t^{-r}\sum_{j=1}^r(-1)^{j+1}\binom{r}{j}j^{-r}\Delta_{jt}^r(f,x)\,.$$

Since $\left\|\Delta_{jt}^r(f)\right\| \le \omega_r(f,jt)\le j^r\omega_r(f,t)$, we obtain

$$t^r\left\|g^{(r)}\right\|_p(A) \le 2^r\omega_r(f,t)_p\,, \tag{2.14}$$

$$\|f-g\|_p(A)+t^r\left\|g^{(r)}\right\|_p(A) \le C\,\omega_r(f,t)_p \tag{2.15}$$

and hence (2.11).

In the case $A=[a,b]$, we also have (2.15) with A replaced by I_1 and for $t\le t_0$. By symmetry, this is also true for the interval $I_2=\left[a+\frac{b-a}{4},b\right]$. Then, Lemma 2.3 yields the right inequality in (2.11) for $t\le t_0$.

Now let g_0 be a function in W_p^r for which (2.15) is valid for $t = t_0$, so that in particular $\|f - g_0\|_p \le C\,\omega_r(f, t_0)_p$, and let P be its Taylor polynomial of degree $r-1$ for the mid-point of A. By the estimate (5.7) of Chapter 2 for the remainder in Taylor's formula,

$$\|g_0 - P\|_p \le |A|^r \left\|g_0^{(r)}\right\|_p \le C\, t_0^r \left\|g_0^{(r)}\right\|_p \le C\,\omega_r(f, t_0)_p .$$

Therefore, writing $f - P = f - g_0 + g_0 - P$, we obtain

$$\|f - P\|_p \le C\,\omega_r(f, t_0)_p \le C\,\omega_r(f,t)_p , \quad t > t_0 .$$

Also $P^{(r)} \equiv 0$. This yields $K(f,t) \le C\,\omega_r(f,t)_p$ for $t \ge t_0$. □

§ 3. Comparisons of Moduli of Smoothness

As an application of the equivalence relation (2.11) of the previous section, we shall compare different moduli of smoothness. Later, in Chapter 12, this will give us information about the relations between smoothness spaces defined using these moduli. The Marchaud inequality (Theorem 8.1 of Chapter 2) is one such result. A similar relation is the following "weak" inverse to the inequality (7.13) of Chapter 2.

Theorem 3.1 (Johnen and Scherer [1976]). *If $A = \mathbb{R}$, $\mathbb{R}_+$, $\mathbb{T}$ or $[a,b]$, there is a constant $C_r > 0$ which depends only on $r = 2, 3, \dots$ with the following property. If $1 \le p \le \infty$ and $f \in L_p(A)$ (or $f \in C(A)$, for $p = \infty$) and $1 \le k < r$, then for $t > 0$*

$$\omega_{r-k}\left(f^{(k)}, t\right)_p \le C_r \int_0^t \frac{\omega_r(f,s)_p}{s^{k+1}}\,ds , \tag{3.1}$$

in the sense that whenever the right side of (3.1) is finite then $f \in W_p^k(A)$ (or $f \in C^k(A)$ if $p = \infty$) and (3.1) holds.

Proof. We fix $t > 0$. When $A = [a,b]$, $\omega_{r-k}\left(f^{(k)}, t\right)_p \le C\,\omega_{r-k}\left(f^{(k)}, |A|\right)_p$ and therefore, we can assume that $t \le |A|$ in this case. Also, if $\omega_r(f, s_0)_p = 0$, then $\omega_r(f,s)_p = 0$, $s \le s_0$, and by Proposition 7.1 of Chapter 2, f is a polynomial of degree $< r$ and both sides of (3.1) are identically zero; we can therefore assume $\omega_r(f,s)_p > 0$ for all $s > 0$ and by (2.11) the same for $K(f,s)$. We use the abbreviation $\lambda_j := 2^{-j}t$. Let g_j, $j = 0, 1, \dots$, satisfy

$$\|f - g_j\|_p + \lambda_j^r \left\|g_j^{(r)}\right\|_p \le 2K\left(f, \lambda_j^r; L_p, W_p^r\right) \le C\,\omega_r(f, \lambda_j)_p . \tag{3.2}$$

Here and later all constants depend only on r. We can write $f = g_0 + \sum_{j=0}^\infty (g_{j+1} - g_j)$ with convergence in L_p. We then apply Theorem 5.6 of Chapter 2 and find from (3.2) that in the L_p norm,

$$\begin{aligned}
\lambda_j^k \left\|g_{j+1}^{(k)} - g_j^{(k)}\right\| &\le C\left\{\|g_{j+1} - g_j\| + \lambda_j^r \left\|g_{j+1}^{(r)} - g_j^{(r)}\right\|\right\} \\
&\le C\left\{\|f - g_j\| + \|f - g_{j+1}\| + \lambda_j^r \left\|g_j^{(r)}\right\| + \lambda_j^r \left\|g_{j+1}^{(r)}\right\|\right\} \\
&\le C\left\{\omega_r(f,\lambda_j) + \omega_r(f,\lambda_{j+1})\right\} \le C\,\omega_r(f,\lambda_j) .
\end{aligned} \tag{3.3}$$

Hence,

$$\sum_{j=0}^\infty \left\|g_{j+1}^{(k)} - g_j^{(k)}\right\| \le C \sum_{j=0}^\infty \frac{\omega_r(f,\lambda_j)}{\lambda_j^k} \le C \int_0^t \frac{\omega_r(f,s)}{s^{k+1}}\,ds . \tag{3.4}$$

Therefore $g_0 + \Sigma_1^\infty [g_{j+1} - g_j]$ converges in W_p^r and since its sum is f, we have $f^{(k)} \in L_p$ $\left(f^{(k)} \in C(A), p = \infty\right)$ and from (3.4)

$$\left\|f^{(k)} - g_0^{(k)}\right\| \le C \int_0^t \frac{\omega_r(f,s)}{s^{k+1}}\,ds .$$

On the other hand, $t^r \left\|g_0^{(r)}\right\| \le C\,\omega_r(f,t)$ by (3.2). Hence, from Theorem 2.4,

$$\begin{aligned}
\omega_{r-k}\left(f^{(k)}, t\right) &\le C\,K\left(f^{(k)}, t^{r-k}; L_p, W_p^{r-k}\right) \\
&\le C\left\{\left\|f^{(k)} - g_0^{(k)}\right\| + t^{r-k}\left\|g_0^{(r)}\right\|\right\} \\
&\le C\left[\int_0^t \frac{\omega_r(f,s)}{s^{k+1}}\,ds + t^{-k}\omega_r(f,t)\right] .
\end{aligned}$$

Since $t^{-k}\omega_r(f,t)$ can be incorporated into the integral, we have (3.1). □

(Note that there is a similar proof of Marchaud's inequality.)

Inequality (3.1) can serve to compare different types of Lipschitz spaces. The following provides a proof of Theorem 9.2 of Chapter 2.

Corollary 3.2. *If $A = \mathbb{R}$, $\mathbb{R}_+$, $\mathbb{T}$, or $[a,b]$, $1 \le p \le \infty$, and $\alpha > 0$, then* (i) *if α is not an integer,*

$$\operatorname{Lip}^*(\alpha, L_p) = \operatorname{Lip}(\alpha, L_p) \tag{3.5}$$

with equivalent seminorms. (ii) *If $\alpha = k+1$ is an integer, then $f \in \operatorname{Lip}^*(\alpha, L_p)$ if and only if $f^{(k)} \in \operatorname{Lip}^*(1, L_p)$ with equivalent seminorms.*

Proof. If α is not an integer, we write $\alpha = k + \beta$, with $0 < \beta < 1$. Let $r := k+1$. From (7.13) of Chapter 2, we have $\omega_r(f,t)_p \le t^k \omega\left(f^{(k)}, t\right)$ and therefore

$$|f|_{\operatorname{Lip}^*(\alpha, L_p)} \le |f|_{\operatorname{Lip}(\alpha, L_p)} . \tag{3.6}$$

On the other hand, if $\omega_r(f,t) \le M t^\alpha$, then Theorem 3.1 shows that $\omega\left(f^{(k)}, t\right)_p \le CMt^\beta$ with C depending only on r and α. Hence,

$$|f|_{\operatorname{Lip}(\alpha, L_p)} \le C\,|f|_{\operatorname{Lip}^*(\alpha, L_p)} .$$

A similar argument with $\alpha = k+1$ and $r := k+2$ proves (ii). □

It is also possible to compare moduli of smoothness for different L_p spaces. For example, if $1 \le p \le q \le \infty$, $r = 1, 2, \dots$, and A is a compact subinterval of $\mathbb{R}$ then by Hölder's inequality, we have

$$\omega_r(f,t)_p \le C\,\omega_r(f,t)_q\,, \quad t > 0\,. \tag{3.7}$$

We shall prove (3.13) below which is a weak inverse of this. We shall use the fact that the l_q norm is smaller than the l_p norm if $q \ge p$:

$$(\Sigma a_i^q)^{1/q} \le (\Sigma a_i^p)^{1/p}\,, \quad a_i \ge 0\,, \quad p \le q\,. \tag{3.8}$$

Lemma 3.3. *Let $A = \mathbb{R}$, $\mathbb{R}_+$, $\mathbb{T}$, or $[a,b]$ and let $\varepsilon_0 := |b-a|$ if $A = [a,b]$ and $\varepsilon_0 := \infty$ otherwise. If $1 \le p \le q \le \infty$, $\theta := 1/p - 1/q$, and $r = 1, 2, \dots$, then there is a constant C depending only on r such that for all $f \in W_p^r$ and all $0 < \varepsilon < \varepsilon_0$,*

$$\omega_r(f,\varepsilon)_q \le C\,\varepsilon^{r-\theta} \left\| f^{(r)} \right\|_p \tag{3.9}$$

$$\|f\|_q \le C\,\varepsilon^{-\theta} \left\{ \|f\|_p + \varepsilon^r \left\| f^{(r)} \right\|_p \right\}. \tag{3.10}$$

Proof. Let $A_c := \{x \in A : x + c \in A\}$. Using Hölder's inequality with (7.10) of Chapter 2, we obtain for $x \in A_{rh}$

$$\begin{aligned} |\Delta_h^r(f,x)| &\le h^r \left\| f^{(r)} \right\|_p [x, x+rh] \left(\int_x^{x+rh} M(u)^{p'}\,du \right)^{1/p'} \\ &\le Ch^{r-1/p} \left\| f^{(r)} \right\|_p [x, x+rh]\,, \quad 1/p + 1/p' = 1\,, \end{aligned} \tag{3.11}$$

because $M(u) \le 1/h$, for all u (this follows from (2.7) of Chapter 5 because $N(u) \le 1$), and M is supported on $[0, rh]$.

Now, we write $A_{rh} = \cup_{I \in \Lambda} I$ as a union of intervals I with disjoint interiors and length $\le \varepsilon$. If $I = [\alpha, \beta]$ is any interval from Λ, and $\widetilde{I} := [\alpha, \beta + r\varepsilon]$, then applying the L_q norm on I to both sides of (3.11), we obtain for $0 \le h \le \varepsilon$,

$$\|\Delta_h^r(f)\|_q\,(I) \le C\,\varepsilon^{r-\theta} \left\| f^{(r)} \right\|_p (\widetilde{I})\,. \tag{3.12}$$

For the partition Λ, we can further require that any point $x \in A_{rh}$ appears in at most $r+2$ of the intervals $\widetilde{I}$. Hence, if we raise both sides of (3.12) to the q-th power and then sum over all $I \in \Lambda$ and apply (3.8), we obtain (3.9).

For the proof of (3.10), we let I be any interval with $\varepsilon/2 \le |I| \le \varepsilon$. To the inequality

$$|I|^{1/p} |f(y)| \le \|f\|_p(I) + |I|\,\|f'\|_p(I)\,, \quad y \in I\,,$$

((5.16) of Chapter 2), we apply an L_q norm over I and we use (5.8) of Chapter 2 to obtain

$$\begin{aligned} \|f\|_q(I) &\le |I|^{-\theta} \left[\|f\|_p(I) + |I|\,\|f'\|_p \right] \\ &\le C\varepsilon^{-\theta} \left[\|f\|_p(I) + \varepsilon^r \left\| f^{(r)} \right\|_p (I) \right]. \end{aligned}$$

We then write A as a disjoint union of such intervals I and use (3.8) to arrive at (3.10). □

Theorem 3.4 (DeVore, Riemenschneider and Sharpley [1979]). *Let $A = \mathbb{R}$, $\mathbb{R}_+$, $\mathbb{T}$ or $[a,b]$ and $r = 1, 2, \dots$. If $1 \le p \le q \le \infty$ and $\theta := 1/p - 1/q$, then there is a constant $C > 0$ depending only on r such that for all $f \in L_p$ and $t > 0$,*

$$\omega_r(f,t)_q \le C \int_0^t s^{-\theta} \omega_r(f,s)_p \frac{ds}{s}\,, \tag{3.13}$$

in the sense that whenever the right side of (3.13) *is finite then $f \in L_q$ and* (3.13) *holds.*

Proof. As in the proof of Theorem 3.1, we can assume that $\omega_r(f,s)_p > 0$ for all s. We again put $\lambda_j := 2^{-j} t$ and select $g_j \in W_p^r$, $j = 0, 1, \dots$, so that

$$\|f - g_j\|_p + \lambda_j^r \left\| g_j^{(r)} \right\|_p \le C\,\omega_r(f, \lambda_j)_p\,. \tag{3.14}$$

Let $h_j := g_{j+1} - g_j$. By the triangle inequality, using (3.14), we find,

$$\|h_j\|_p + \lambda_j^r \left\| h_j^{(r)} \right\|_p \le C\,\omega_r(f, \lambda_j)_p\,, \quad j = 1, 2, \dots\,. \tag{3.15}$$

Now suppose that the integral in (3.13) is finite for $t > 0$ with $t \le |A|$ if $A = [a,b]$. It follows from (3.15) that $f = g_0 + \sum_{j=0}^{\infty} h_j$ in the sense of convergence in L_p. We also have from (3.10) and (3.15)

$$\begin{aligned} \sum_{j=0}^{\infty} \|h_j\|_q &\le C \sum_{j=1}^{\infty} \lambda_j^{-\theta} \left[\|h_j\|_p + \lambda_j^r \left\| h_j^{(r)} \right\|_p \right] \\ &\le C \sum_{j=1}^{\infty} \lambda_j^{-\theta} \omega_r(f, \lambda_j)_p \le C \int_0^t s^{-\theta} \omega_r(f,s)_p \frac{ds}{s}\,. \end{aligned} \tag{3.16}$$

From (3.10), g_0 is in L_q and hence $f \in L_q$. We use (3.9), (3.14), and (3.16) to obtain

$$\begin{aligned} \omega_r(f,t)_q &\le \omega_r(g_0,t)_q + \sum_{j=1}^{\infty} \omega_r(h_j,t)_q \\ &\le C \left[t^{r-\theta} \left\| g_0^{(r)} \right\|_p + \sum_{j=0}^{\infty} \|h_j\|_q \right] \\ &\le C \left[t^{-\theta} \omega_r(f,t)_p + \int_0^t s^{-\theta} \omega_r(f,s)_p \frac{ds}{s} \right]. \end{aligned}$$

Since the term $t^{-\theta}\omega_r(f,t)_p$ can be incorporated into the integral, we obtain (3.13) in all cases, provided $t \le |A|$ if $A = [a,b]$. If $A = [a,b]$ and $t > |A|$, then $\omega_r(f,t)_q = \omega_r(f,|A|)_q$ and (3.13) follows in full generality. □

Theorem 3.4 carries with it information about the spaces $\mathrm{Lip}^*(\alpha, L_p)$ as p varies. For example, if $\omega_r(f,t)_p = \mathcal{O}(t^{\alpha+\theta})$, then from (3.13), $\omega_r(f,t)_q = \mathcal{O}(t^\alpha)$. From this, we obtain the continuous embeddings

$$(3.17) \qquad \mathrm{Lip}^*(\alpha+\theta, L_p) \subset \mathrm{Lip}^*(\alpha, L_q), \quad \alpha > 0.$$

It is easy to see that θ cannot be replaced by a smaller number in (3.17). For example, if f is the characteristic function χ_J of a proper subinterval J of A, then for $0 < s < \infty$, $\omega(f,t)_s \sim t^{1/s}$.

We have similar embeddings for Besov spaces. For example, from the Hardy inequality (3.1) of Chapter 2 (and Theorem 3.5 of Chapter 2 in the case $q < 1$) applied to (3.13), we have for any $\alpha > 0$, $0 < q \le \infty$ the continuous embeddings

$$(3.18) \qquad B_q^{\alpha+\theta}(L_p) \subset B_q^\alpha(L_q).$$

For the limiting case $\alpha = 0$, we have already shown in the proof of Theorem 3.4 that

$$(3.19) \qquad B_1^\theta(L_p) \subset L_q.$$

Later, in Theorem 8.1 of Chapter 11, we shall see that this last embedding can be improved by replacing 1 by p.

§ 4. Two Theorems of Whitney

The simple proofs below of two important theorems of Whitney are examples of applications of K-functionals. It is easy to prove that a function $f \in W_p^r(I)$, $I = [a,b] \subset \mathbb{R}$ can be extended to the whole line with the norm of the extended function f_1 satisfying $\|f_1\|_{W_p^r(\mathbb{R})} \le C\|f\|_{W_p^r(I)}$ where C is a constant. A deeper result is Whitney's theorem [1934], that an extension can be achieved with the preservation of the order of magnitude of $\omega_r(f,t)_p$.

Theorem 4.1. *Let $I \subset J$ be two compact intervals and let $1 \le p \le \infty$. Then there is a bounded linear operator T which maps $L_p(I)$ into $L_p(J)$, $1 \le p < \infty$, $C(I)$ into $C(J)$ for $p = \infty$, which is an extension: $Tf(x) = f(x)$, $x \in I$, and which preserves polynomials of degree $< r$. Furthermore, there is a constant C depending only on r and $|J|/|I|$, for which*

$$(4.1) \qquad \omega_r(Tf,t)_p(J) \le C\,\omega_r(f,t)_p(I), \quad t \ge 0.$$

Proof. It is sufficient to prove this when $I = [0,1]$, $J = [-1,1]$, for the general case follows by symmetry, translations, and iterations. We put

$$(4.2) \qquad Tf(x) := \begin{cases} f(x), & 0 \le x \le 1 \\ \sum_{i=0}^{r-1} \alpha_i f(-2^{-i}x), & -1 \le x < 0 \end{cases}$$

with constants α_i that satisfy

$$(4.3) \qquad \sum_{i=0}^{r-1} \alpha_i (-2^{-i})^j = 1, \quad j = 0,1,\dots,r-1.$$

The constants α_i exist since the determinant of the system (4.3) is a Vandermonde determinant.

It is clear that T is a bounded operator from $L_p(I)$ into $L_p(J)$ (or from $C(I)$ into $C(J)$) with norm depending only on r. Condition (4.3) guarantees that polynomials of degree $< r$ are preserved by the extension.

The operator T also maps $W_p^r(I)$ into $W_p^r(J)$. Indeed, if $g \in W_p^r(I)$, then from (4.3) the derivatives $(Tg)^{(j)}$, $j = 0,1,\dots,r-1$ are continuous at $x = 0$, hence they must be absolutely continuous on J. Also, $\|(Tg)^{(r)}\|_p \le C\|g^{(r)}\|_p$. For $f \in L_p(I)$, $1 \le p < \infty$ or $f \in C(I)$, $p = \infty$, Theorem 2.4 and (1.14) give

$$\begin{aligned}\omega_r(Tf,t)_p(J) &\le C\,K\left(Tf,t;L_p(J),W_p^r(J)\right) \le C\,K\left(f,t;L_p(I),W_p^r(I)\right)\\ &\le C\,\omega_r(f,t)_p(I), \quad t > 0. \qquad \square\end{aligned}$$

Example. Each function $f \in C[0,1]$ possesses an even extension Tf onto $[-1,1]$ with the property

$$(4.4) \qquad \omega(Tf,h) \le C\omega(f,h), \quad h > 0.$$

This is the case $r = 1$ of Theorem 4.1.

Another theorem of Whitney [1957] deals with the approximation of a function on an interval I by polynomials of degree $< r$, with the error $E_{r-1}(f,I)_p$ estimated in terms of the length of I. One obtains good approximation, for fixed r, if $|I|$ is small. This is in contrast with the more usual theorems of Jackson type (see Chapter 7) where I is fixed but the degree of the polynomials is large.

Theorem 4.2. *For each $r = 1,2,\dots$, there is a constant $C_r > 0$, such that for each $f \in L_p(I)$, $1 \le p \le \infty$ (with $L_\infty(I)$ replaced by $C(I)$),*

$$(4.5) \qquad C_r E_{r-1}(f,I)_p \le \omega_r(f,|I|)_p \le 2^r E_{r-1}(f,I)_p.$$

Remark. As will be shown in Chapter 12, this theorem also holds for $p < 1$.

Proof. Let $g \in W_p^r(I)$ be arbitrary and let P be the Taylor polynomial for g of degree $r-1$ corresponding to one of the endpoints of I. By (5.7) of Chapter 2, we have in the $L_p(I)$ norm

$$\|g - P\| \le \frac{1}{(r-1)!}|I|^r\|g^{(r)}\|.$$

Hence,

$$\|f-P\| \le \|f-g\| + |I|^r\|g^{(r)}\| .$$

We now take an infimum over all g to obtain $E_r(f,I) \le K(f,|I|^r;L_p,W_p^r(I))$. Hence, the first inequality of (4.5) follows from Theorem 2.4. On the other hand, if P is any polynomial of degree $< r$, then $\omega_r(f,|I|)_p = \omega_r(f-P,|I|)_p \le 2^r\|f-P\|_p$. Taking for P a best $L_p(I)$ approximation to f from $\mathcal{P}_{r-1}$, we obtain the right inequality (4.4). □

It is of interest to know the dependence of the constant C_r in (4.5) on r. Whitney's proof of Theorem 4.2 (for $p=\infty$) establishes only the existence of constants C_r for which (4.5) is valid. Our proof uses Theorem 2.4 and gives only that C_r^{-1} grows at most exponentially. However, Sendov [1987] for the case $p=\infty$, showed that the C_r can be made independent of r. He even showed that one can take $C_r = 1/6$, $r=0,1,\dots$, if $I=[-1,1]$.

§ 5. Averaged Moduli of Smoothness

The modulus of smoothness $\omega_r(f,t)_p = \sup_{0<s\le t}\|\Delta_s^r(r,\cdot)\|_p(A)$ in the Whitney estimate (4.5) is sometimes not suitable for applications because it is not easy to add up several such estimates for different intervals. We therefore introduce some new moduli of smoothness obtained by averaging. These will be applied later in this chapter (in the proof of Theorem 6.2) and to spline approximation in Chapter 12. The simplest such modulus for a function f on $I:=[a,b]$ is given by

$$\text{(5.1)}\qquad \mathrm{w}_r(f,t)_p := \frac{1}{t}\int_0^t \|\Delta_s^r(f,\cdot)\|_p(I_{rs})ds\,, \quad 0\le p\le\infty$$

where as usual $I_c := [a,b-c]$.

It is also useful to allow the averaging to take place on other intervals. If $0\le\alpha<\beta\le 1$ are fixed, we let

$$\text{(5.2)}\qquad \mathrm{w}_r^*(f,t)_p := \frac{1}{t}\int_{\alpha t}^{\beta t}\|\Delta_s^r(f,\cdot)\|_p(I_{rs})ds\,.$$

We shall show that w_r and w_r^* are equivalent to ω_r. For this purpose, we shall use the following identity for differences:

$$\text{(5.3)}\qquad \Delta_h^r(f,x) = \sum_{k=1}^{r}(-1)^k\binom{r}{k}\left[\Delta_{ks}^r(f,x+kh) - \Delta_{h+ks}^r(f,x)\right]\,,$$

which holds for every $s\in\mathbb{R}$. To prove (5.3), we note that

$$\begin{aligned}
&\sum_{k=0}^{r}(-1)^{r+k}\binom{r}{k}\Delta_{h+ks}^r(f,x)\\
&\quad= \sum_{k=0}^{r}(-1)^{r+k}\binom{r}{k}\sum_{j=0}^{r}(-1)^{r+j}\binom{r}{j}f(x+j(h+ks))\\
&\quad= \sum_{j=0}^{r}(-1)^{r+j}\binom{r}{j}\sum_{k=0}^{r}(-1)^{r+k}\binom{r}{k}f(x+jh+ksj)\\
&\quad= \sum_{j=1}^{r}(-1)^{r+j}\binom{r}{j}\Delta_{js}^r(f,x+jh)
\end{aligned}$$

because the term corresponding to $j=0$ is 0. Moving all the terms except $k=0$ of the left side of (5.4) to the right side yields (5.3).

In the following lemma which compares averaged moduli with ω_r, we shall assume (purely to avoid technicalities) that $\beta - r\alpha - (2r)^{-1} > 0$.

Lemma 5.1. *If $0\le\alpha<\beta\le 1$ and $\beta-r\alpha-(2r)^{-1}>0$, then for all $f\in L_p(I)$, $1\le p<\infty$ (or $f\in C(I)$ if $p=\infty$), we have*

$$\text{(5.5)}\qquad C_1\mathrm{w}_r^*(f,t)_p \le \omega_r(f,t)_p \le C_2\mathrm{w}_r^*(f,t)_p\,,\quad 0<t\le t_0 := |I|(4r)^{-1}$$

where C_1, C_2 are positive constants depending only on r and $\beta-r\alpha-(2r)^{-1}$. In particular,

$$\text{(5.6)}\qquad C_1\mathrm{w}_r(f,t)_p \le \omega_r(f,t)_p \le C_2\mathrm{w}_r(f,t)_p\,,\quad 0<t\le t_0\,.$$

Proof. It is enough to prove (5.5). The left inequality in (5.5) is immediate because $\|\Delta_s^r(f)\|_p(I_{rs}) \le \omega_r(f,t)_p$, for $0\le s\le t$. For the right inequality, we note that $t\le t_0/r$, $0\le h\le t/2$ and $0\le s\le t$ imply that $rkt+kh\le |I|/2$, so that $x\in I_- := [a,(a+b)/2]$ implies $x+kh\in I_{rks}$ and $x\in I_{r(h+ks)}$. Therefore, applying the L_p norm in (5.3), we obtain

$$\text{(5.7)}\quad \|\Delta_h^r(f)\|\,(I_-) \le \sum_{k=1}^{r}\binom{r}{k}\left[\|\Delta_{ks}^r(f)\|\,(I_{rks}) + \left\|\Delta_{h+ks}^r(f)\right\|\left(I_{r(h+ks)}\right)\right]\,.$$

We now take an average in (5.7) with respect to $s\in[\alpha' t,\beta' t]$ with $\alpha' := \alpha r$ and $\beta' := \beta - \frac{1}{2r}$. A change of variables gives

$$\text{(5.8)}\qquad \begin{aligned}
\|\Delta_h^r(f)\|\,(I_-) &\le \frac{C}{t}\sum_{k=1}^{r}\left(\int_{\alpha' kt}^{\beta' kt}\|\Delta_u^r(f)\|(I_{ru})du + \int_{h+\alpha' kt}^{h+\beta' kt}\|\Delta_u^r(f)\|(I_{ru})du\right)\\
&\le \frac{C}{t}\int_{\alpha rt}^{\beta rt}\|\Delta_u^r(f)\|(I_{ru})du = C\mathrm{w}_r^*(f,rt)\,.
\end{aligned}$$

By symmetry, we have that $\|\Delta_{-h}^r(f)\|(I_+)$, $I_+ := [(a+b)/2,b]$, also does not exceed the right side of (5.8). Since this is true for all $0<h\le t/2$, the same follows for $\omega_r(f,t/2)$. We therefore obtain

$$\omega_r(f,rt) \le C\omega_r(f,t/2) \le C\mathrm{w}_r^*(f,rt)\,,\qquad 0\le t\le t_0/r\,. \qquad □$$

§ 6. Moduli of Smoothness with Weights

We shall study here a modification of the moduli of smoothness based on differences Δ_u^r in which the step is of the form $u = \phi(x)h$ and is therefore allowed to depend upon the position of x in the interval $[-1,1]$. The motivation for this lies in the properties of algebraic polynomial approximation. The requirements on the smoothness of f at x can be relaxed if x is close to ± 1 without impairing the error of approximation.

Several authors, among them Ivanov [1980], have introduced moduli of this type, but the most useful proved to be the latest version, given in the monograph of Ditzian and Totik [A-1987], based on their earlier papers. We give here, in §8 of Chapter 8 and in §7 of Chapter 10, a simplified exposition of this theory for the most important case of the weight $\phi(x) = (1-x^2)^{\frac{1}{2}}$ on the interval $A := [-1,1]$.

We shall use symmetric differences $\widetilde{\Delta}_h^r(f,x) := \Delta_h^r(f, x - rh/2)$ on A, and we shall adopt the convention that *the difference is zero* if one of the values of f needed for its evaluation is at a point outside of A. Thus $\widetilde{\Delta}_h^r(f,x)$ is zero if x is outside of $[-1+rh/2, 1-rh/2]$. The weighted moduli of smoothness of order $r = 1, 2, \dots$ are defined for $f \in L_p$, $0 < p \le \infty$, by means of

$$\omega_r^\phi(f,t)_p := \sup_{0<h\le t} \left\| \widetilde{\Delta}_{h\phi(\cdot)}^r(f,\cdot) \right\|_p, \quad \phi(x) = \sqrt{1-x^2}. \tag{6.1}$$

According to our convention, $\widetilde{\Delta}_{h\phi(x)}^r(f,x) = 0$ unless $x \pm rh\phi(x)/2 \in A$, or equivalently, $x \in A(h) := [-a_h, a_h]$ with $a_h := \frac{1-(rh/2)^2}{1+(rh/2)^2}$. Clearly, $A(h) \subset A(h')$, if $h' < h$.

A useful fact is that any function $\psi(x) := x + s\phi(x)$ with $|s| \le rh/2$ satisfies

$$\psi'(x) \ge \tfrac{1}{2}, \quad x \in A(h). \tag{6.2}$$

In fact, for $x \in A(h)$, the function $|s\phi'(x)|$ takes its largest value when $x = \pm a_h$, in which case $|s\phi'(a_h)| = \frac{|s|}{rh}\left[1-(rh/2)^2\right] \le \frac{1}{2}$.

From (6.2), we can deduce that $\omega_r^\phi(f,t)_p$ *is defined and finite for every* $f \in L_p(A)$, $0 < p \le \infty$, *and* $t \ge 0$. In fact, $\|\widetilde{\Delta}_{h\phi(\cdot)}^r(f,\cdot)\|_p^p(A(h))$ is less than a constant times the sum of integrals $J_k := \int_{A(h)} |f(x + kh\phi(x))|^p dx$, $k = -r/2, -r/2+1, \dots, r/2$. By (6.2),

$$J_k \le 2\int_{A(h)} |f(x + kh\phi(x)|^p \left(1 + kh\phi'(x)\right) dx \le 2\|f\|_p^p. \tag{6.3}$$

Therefore,

$$\omega_r^\phi(f,t)_p \le C\|f\|_p, \quad t > 0. \tag{6.4}$$

Our main result (Theorem 6.2 below) states that ω_r^ϕ is equivalent to the following modified K-functional. We let $W_p^r(\phi)$, $1 \le p \le \infty$, $r = 1, 2, \dots,$

denote the set of all f on A for which $f^{(r-1)}$ is absolutely continuous on $(-1,1)$ and

$$|f|_{W_p^r(\phi)} := \left\|\phi^r f^{(r)}\right\|_p (A) \tag{6.5}$$

is finite. If $f \in L_p(A)$, we have by (1.10)

$$K\left(f,t;L_p,W_p^r(\phi)\right) := \inf_g \left\{ \|f-g\|_p(A) + t|g|_{W_p^r(\phi)} \right\}, \tag{6.6}$$

where the infimum is over all $g \in W_p^r(\phi)$.

The equivalence of K with ω_r^ϕ will be established only for sufficiently small t. Our proof is simplest when $t \le (2r)^{-1}$, which we will assume throughout. We begin with the following analogue of (7.12) of Chapter 2.

Theorem 6.1 (Ditzian and Totik [A-1987]). *There is a constant $C > 0$ depending only on r such that for each $f \in W_p^r(\phi)$, $1 \le p \le \infty$, we have*

$$\omega_r^\phi(f,t)_p \le C\,t^r |f|_{W_p^r(\phi)}, \quad 0 \le t \le (2r)^{-1}. \tag{6.7}$$

Proof. Suppose $0 < h \le (2r)^{-1}$. We let $\lambda := \lambda(x) := rh\phi(x)/2$ and denote by $\chi(x,u)$ the characteristic function on $\mathbb{R}^2$ of the set $E := \{(x,u) : x \in A(h), u \in A, |x-u| \le \lambda(x)\}$. We shall use the following inequalities:

$$\text{(6.8)}\qquad \begin{aligned} &\text{(i)} \quad \int_{-\infty}^{\infty} \chi(x,u)dx \le 2rh\phi(u), \quad u \in A \\ &\text{(ii)} \quad \int_{-\infty}^{\infty} \frac{\chi(x,u)}{\phi(u)}du \le 8rh, \quad x \in A(h). \end{aligned}$$

For the evaluation of the integral in (i), we can by symmetry assume that $u \ge 0$. We solve the equation $|u-x| = \lambda(x)$ for x and find $x = [u \pm \eta(1-u^2 + \eta^2)^{\frac{1}{2}}]/(1+\eta^2) =: b_\pm$, with $\eta := rh/2 \le 1/4$. The integral in (i) is then the length of the interval $[b_-, b_+] \cap A(h)$. If $1-u^2 \ge \eta^2$, this integral is less than $2\eta(1-u^2+\eta^2)^{\frac{1}{2}} \le 2rh\phi(u)$. On the other hand, if $1-u^2 \le \eta^2$, this integral does not exceed $a_h - b_- \le 1 - u + \eta(1-u^2+\eta^2)^{\frac{1}{2}} - \eta^2$. Since

$$\eta(1-u^2+\eta^2)^{\frac{1}{2}} - \eta^2 = \eta\frac{1-u^2}{(1-u^2+\eta^2)^{\frac{1}{2}} + \eta} \le 1-u^2,$$

we have $a_h - b_- \le 2(1-u^2) \le rh\phi(u)$.

To prove (ii), by symmetry, we can assume that $x \ge 0$. From our assumption $h \le (2r)^{-1}$, we get $u \ge -rh/2 \ge -1/4$ whenever $\chi(x,u) > 0$. Therefore $\phi(u) \ge \frac{1}{2}(1-u)^{\frac{1}{2}}$ under the same condition. Now, when $x \in A(h)$, we have $\lambda(x) \le 1-x$. Therefore, the integral in (ii) is less than or equal to

$$\begin{aligned} &4\left[(1-x+\lambda(x))^{\frac{1}{2}} - (1-x-\lambda(x))^{\frac{1}{2}}\right] \\ &\qquad = 8\lambda(x)\left[(1-x+\lambda(x))^{\frac{1}{2}} + (1-x-\lambda(x))^{\frac{1}{2}}\right]^{-1} \end{aligned}$$

and this does not exceed $4rh(1-x^2)^{\frac{1}{2}}/(1-x)^{\frac{1}{2}} \le 8rh$, as desired.

To prove (6.7) for $r = 1$, we write

$$\left|\frac{1}{h}\widetilde{\Delta}_{h\phi(x)}(f,x)\right| \le \int_{-\infty}^{\infty} |f'(u)|\frac{\chi(x,u)}{h}du = \int_{-\infty}^{\infty} |f'(u)|\phi(u)K(x,u)du\,,$$

where $K(x,u) := \frac{\chi(x,u)}{h\phi(u)}$. Since by (i) and (ii), the integral of K with respect to x and to u over $(-\infty,\infty)$ is $\le 8r = 8$, we can apply Theorem 4.5 of Chapter 2. This yields $\omega_r^{\phi}(f,h)_p \le Ch\|f'\phi\|_p$, that is, (6.7).

To prove (6.7) for $r > 1$, we let $M(u,x) := \lambda^{-1}M(u/\lambda)$ where $\lambda := \lambda(x) = rh\phi(x)/2$, as above, and M is the B-spline of order r with knots -1, $-1+2/r,\dots,1$. If $\widetilde{\Delta}^r_{h\phi(x)}(f,x) \ne 0$, then from (2.2) of Chapter 5, we have

$$\left|\widetilde{\Delta}^r_{h\phi(x)}(f,x)\right| \le (h\phi(x))^r \int_{-\lambda}^{\lambda}\left|f^{(r)}(x+u)\right| M(u,x)du$$

$$\le \frac{Bh^r}{\lambda}\int_{-\lambda}^{\lambda}\left|f^{(r)}(x+u)\right|\phi(x+u)^r du \tag{6.9}$$

$$= Bh^r\int_{-1}^{1}\left|f^{(r)}(x+\lambda u)\right|\phi(x+\lambda u)^r du$$

with

$$B := \sup_{|u|<\lambda, x\in A(h)} \Phi(u,x)\,; \quad \Phi(u,x) := \lambda M(u,x)\left(\frac{\phi(x)}{\phi(x+u)}\right)^r .$$

We next show that B is finite. Let $x \ge 0$ with $x \in A(h)$. Then, $u \ge -\lambda(x) \ge -1/4$ and so $\phi(x+u) \ge \frac{1}{2}(1-x-u)^{1/2}$. Also, since M has a zero of order $r-1$ at ±1 (see (2.6)(ii) of Chapter 5) we have $M(u) \le C(1-u)^{r-1} \le C(1-u)^{r/2}$, $|u| \le 1$. Therefore,

$$\Phi(u,x) \le C\left(\frac{(1-x)(1-u/\lambda)}{1-x-u}\right)^{r/2} \le C\,, \quad |u| \le \lambda\,, \tag{6.10}$$

because the function $(1-u/\lambda)/(1-x-u)$ is decreasing for $|u| \le \lambda$ (since $\lambda \le 1-x$, $x \in A(h)$) and hence has maximum $2/(1-x+\lambda) \le 2/(1-x)$. By symmetry, (6.10) also holds when $x \le 0$. Thus B is finite.

If we apply the $L_p(A_h)$ norm to (6.4) and use (6.2) (in the same way as in (6.3)), we find that

$$\left\|\widetilde{\Delta}^r_{h\phi(\cdot)}(f,\cdot)\right\|_p (A_h) \le 2Bh^r\left\|\phi^r f^{(r)}\right\|_p (A)$$

and (6.7) follows. □

The following result of Ditzian and Totik [A-1987] establishes the equivalence of weighted K-functionals and moduli of smoothness.

Theorem 6.2. *For $1 \le p \le \infty$ and $r = 1,2,\dots$, there are constants C_1, $C_2 > 0$, which depend only on r, such that for all $f \in L_p$*

$$C_1\omega_r^{\phi}(f,t)_p \le K\left(f,t^r;L_p,W_p^r(\phi)\right) \le C_2\omega_r^{\phi}(f,t)_p\,, \quad 0 < t \le (2r)^{-1}. \tag{6.11}$$

Proof. (a) The lower estimate in (6.11) follows easily. For any $f \in L_p$ and $g \in W_p^r(\phi)$, we have from (6.4) and (6.7),

$$\omega_r^{\phi}(f,t)_p \le \omega_r^{\phi}(f-g,t)_p + \omega_r^{\phi}(g,t) \le C\left\{\|f-g\|_p + t^r\|\phi^r g^{(r)}\|_p\right\}.$$

If we take an infimum over all $g \in W_p^r(\phi)$, we obtain the result.

(b) To obtain the upper estimate in (6.11), we shall approximate $f \in L_p(A)$, $A = [-1,1]$, by $Q_T(f)$ where Q_T is a quasi-interpolant spline operator of order $r+1$ with knots T described below. It is sufficient to establish (6.11) when $t = m^{-1}$ and $m \ge 2r$ is an integer; for other values of t, the estimate will follow from this due to the monotonicity of K and ω_r^{ϕ}, and the subadditivity of K with respect to t. We assume that $p < \infty$; the case $p = \infty$ follows with obvious modifications.

We want the knots $T := (t_j)_{-m+1}^{m-1}$ to have Δt_j and $t\phi(t_j)$ of the same order: this can be accomplished if we take $t_j := 1-(m-j)^2t^2$, $j = 0,\dots,m$, $t_j := 1$, $j > m$, and $t_{-j} := -t_j$, $j > 0$. With these choices,

$$t_{j+1}-t_j \le 4t\phi\left(t_{j+1}\right) \le 4t\phi(t_j) \le 8\left(t_{j+1}-t_j\right)\,, \quad j = 0,\dots,m-2\,. \tag{6.12}$$

From this, we have in particular

$$\phi(t_j) \le 8\phi\left(t_{j+1}\right)\,, \quad 0 \le j \le m-2\,, \tag{6.13}$$

and

$$t_{j+1}-t_j \le 8\left(t_{j+2}-t_{j+1}\right)\,, \quad 0 \le j \le m-2\,. \tag{6.14}$$

For $Q := Q_T$, we have according to (4.17) of Chapter 5,

$$\|f-Q(f)\|^p \le C\sum_{j=-m}^{m-1} E_r(f,I_j)^p \le C\sum_{j=-m}^{m-1} E_{r-1}(f,I_j)^p \tag{6.15}$$

where $E_{r-1}(f,I_j)$ is the error of approximation of the function f in the L_p norm on the interval $I_j := [t_{j-r},t_{j+r+1}]$ by algebraic polynomials of degree $r-1$. (Here and it what follows, $C, C_1, \dots$ are constants > 0, which depend only on r.)

We estimate the j-th term of the sum for $j = 0,1,\dots,m-1$. For $J := [a,b] := I_j$, we put

$$\xi := \phi(t_j)\,. \tag{6.16}$$

Then,

$$\frac{1}{2}t\xi \le |J| \le 8^{r+3}t\xi\,. \tag{6.17}$$

Indeed, from (6.12), $t\xi \le 2\left(t_{j+1}-t_j\right) \le 2|J|$ which is the lower inequality in (6.17), while from (6.14) and the symmetry of the t_j (with respect to 0),

$$|J| = (t_{j+r+1}-t_j) + (t_j - t_{j-r}) \le 3\left(t_j - t_{j-r}\right) \le 3(8+\dots+8^r)\left(t_{j+1}-t_j\right)$$

and therefore the right inequality in (6.17) follows from (6.12).

We use Whitney's estimate (4.5) and the averaged modulus of smoothness w_r^* of (5.2) for $\alpha := (4r)^{-1}$, $\beta := 1$. We let $\eta := (8^{r+3}r)^{-1}t\xi$. Then $\eta \le |J|/(4r)$ and we can apply the equivalence (5.5) of ω_r and w_r^* to obtain

$$E_{r-1}(f,J) \le C\omega_r(f,|J|)(J) \le C\omega_r(f,\eta)(J) \le C\mathrm{w}_r^*(f,\eta)$$
$$= \frac{C}{\eta}\int_{\alpha\eta}^{\eta}\left(\int_{a+\frac{rs}{2}}^{b-\frac{rs}{2}}\left|\widetilde{\Delta}_s^r(f,x)\right|^p dx\right)^{1/p} ds\,.$$

From Hölder's inequality, we find

$$(6.18)\qquad \begin{aligned} E_{r-1}(f,J)^p &\le \frac{C}{\eta}\int_{\alpha\eta}^{\eta}\left(\int_{a+\frac{rs}{2}}^{b-\frac{rs}{2}}\left|\widetilde{\Delta}_s^r(f,x)\right|^p dx\right) ds \\ &\le \frac{C}{t}\int_A\left(\frac{1}{\xi}\int_{\alpha\eta}^{\eta}\left|\widetilde{\Delta}_s^r(f,x)\right|^p \chi_J(x,s)ds\right)dx\,. \end{aligned}$$

Here the function χ_J is defined by $\chi_J(x,s) := 1$ if x,s satisfy $\left[x-\frac{1}{2}rs, x+\frac{1}{2}rs\right] \subset J$, and $\chi_J(x,s) := 0$ otherwise.

We would like to estimate the interior integral in the last part of (6.18). We shall use the inequalities

$$(6.19)\qquad \eta t^{-1} \le \phi(x) \le C_2\xi\,, \qquad \text{if } \chi_J(x,s) \ne 0\,, \quad s \ge \alpha\eta\,.$$

Indeed, the maximum of $\phi(x)$, $x \in J$ is assumed at t_{j-r} if $t_{j-r} \ge 0$ or at 0 if $t_{j-r} \le 0$ and therefore the right inequality follows from (6.13) and the symmetry of the points t_k. To prove the left inequality we consider the following three cases. If $x \le t_{m-1}$, then $\phi(x) \ge 8^{-r}\phi(t_j)$ because of (6.13). Then the left inequality of (6.19) follows from the definition of η. If $t_{m-1} < x \le 1 - t^2/64$, then $\phi(x) \ge t/8 \ge \phi(t_{m-1})/8 \ge 8^{-r-1}\phi(t_j)$ and again the left inequality of (6.19) follows from the definition of η. Finally, if $x \ge 1 - t^2/64$, then $8(1-x) \le t\sqrt{1-x} \le t\phi(x)$ and again the left inequality of (6.19) follows from $\eta \le s/\alpha = 8rs/2 \le 8(1-x)$ because $x \le 1 - rs/2$.

From (6.19), we obtain

$$(6.20)\qquad \begin{aligned} \frac{1}{\xi}\int_0^{\eta}\left|\widetilde{\Delta}_s^r(f,x)\right|^p \chi_J(x,s)ds &\le \frac{C}{\phi(x)}\int_0^{t\phi(x)}\left|\widetilde{\Delta}_s^r(f,x)\right|^p \chi_J(x,s)ds \\ &= C\int_0^t\left|\widetilde{\Delta}_{s\phi(x)}^r(f,x)\right|^p \chi_J(x,s\phi(x))ds\,. \end{aligned}$$

This establishes that for any $J = I_j$, $j = 0,\dots,m-1$

$$(6.21)\qquad E_{r-1}(f,J)^p \le \frac{C}{t}\int_0^t\int_A\left|\widetilde{\Delta}_{s\phi(x)}^r(f,x)\right|^p \chi_J(x,s\phi(x))dx\,ds\,.$$

In the same way, we show that (6.21) holds for $j = -m,\dots,-1$.

Now, a point $y \in A$ appears in at most $2r+1$ of the intervals I_j. Hence, $\sum_{j=-m}^{m-1}\chi_{I_j}(x,s) \le (2r+1)\chi_A(x,s)$. Using this and (6.21) in (6.15) gives

$$(6.22)\qquad \|f - Q(f)\|^p \le \frac{C}{t}\int_0^t\int_A\left|\widetilde{\Delta}_{s\phi(x)}^r(f,x)\right|^p \chi_A(x,s)dx\,ds \le C\omega_r^\phi(f,t)^p\,.$$

We can also estimate the derivative of $Q(f)$. From (4.19) of Chapter 5, we have for $x \in (t_j,t_{j+1})$, $j = 0,\dots,m-1$,

$$(6.23)\qquad \left|Q(f)^{(r)}(x)\right| \le C\,(t_{j+1}-t_j)^{-r-1/p}\,E_{r-1}(f,I_j)_p\,.$$

Now $t\phi(x) \le 2\,(t_{j+1}-t_j)$, if $x \in (t_j,t_{j+1})$, because of (6.12). Therefore,

$$t^r\left\|\phi^rQ(f)^{(r)}\right\|_p(t_j,t_{j+1}) \le CE_{r-1}(f,I_j)_p\,, \quad j = 0,\dots,m-1\,.$$

Similarly, for $j = -m,\dots,-1$. We add these estimates, using (6.21), as above and obtain

$$(6.24)\qquad t^r\left\|\phi^rQ(f)^{(r)}\right\|_p \le C\omega_r^\phi(f,t)_p\,.$$

The upper estimate in (6.11) now follows from (6.22), (6.24) and the definition of the K-functional. □

Here is a simple application of Theorem 6.2. Since W_p^r is continuously embedded in $W_p^r(\phi)$, any function $f \in L_p(I)$ can be approximated arbitrarily closely by functions $g \in W_p^r(\phi)$. Hence, the K-functional (6.6) tends to 0 as $t \to 0$. In view of (6.11), we also have $\omega_r^\phi(f,t)_p \to 0$.

§ 7. The θ,q-Interpolation Spaces

Let (X_0,X_1) be a *pair* of linear quasinormed (or normed) complete spaces of §1, embedded in a linear Hausdorff topological space $\mathfrak{X}$. Then the spaces $X_0 \cap X_1$, $X_0 + X_1$ with quasinorms (1.4), (1.5) are well defined. We call a third quasinormed space X an *intermediate space* for the pair (X_0,X_1) if there are continuous embeddings

$$(7.1)\qquad X_0 \cap X_1 \subset X \subset X_0 + X_1\,.$$

If $X_0 \subset X_1$, then (7.1) simply means that $X_0 \subset X \subset X_1$. It is easy to prove that each rearrangement-invariant space on $\mathbb{R}$ or its subinterval is intermediate between L_∞ and L_1 (see Bennett and Sharpley [B-1988, p. 78]).

The present section is devoted to one of the most important types of intermediate spaces. Their construction uses the K-functional of the pair (X_0,X_1) and the θ,q-quasinorms ($\theta > 0, 0 < q \le \infty$) of §3, Chapter 2.

For a given pair (X_0,X_1) we put

$$(7.2)\qquad \rho(f)_{\theta,q} := \|K(f,\cdot)\|_{\theta,q} := \begin{cases} \left(\displaystyle\int_0^\infty \left[t^{-\theta}K(f,t)\right]^q \frac{dt}{t}\right)^{1/q}, & 0 < q < \infty \\ \displaystyle\sup_{0\le t<\infty} t^{-\theta}K(f,t)\,, & q = \infty\,. \end{cases}$$

The space $X_{\theta,q} := (X_0, X_1)_{\theta,q}$ consists of all functions $f \in X_0 + X_1$ for which (7.2) is finite. From the proof of Theorem 3.2 of Chapter 2 we derive that $X_{\theta,q}$ is a quasinormed space.

There is another case when this construction works. Let K be defined by (1.10) with the help of the seminorm $|f|_{X_1}$. Then the expression on the right in (7.2), which we still denote $\rho(f)_{\theta,q}$, is only a quasi-seminorm. In this case we shall always assume that the null space $\mathcal{N}$ of $|\cdot|_{X_1}$ is contained in X_0 and that the following embedding inequality holds:

$$\operatorname{dist}(f, \mathcal{N}) := \inf_{g \in \mathcal{N}} \|f - g\|_{X_0} \le C|f|_{X_1}\,. \tag{7.3}$$

Then $X_0 = X_0 + X_1$ and $\|\cdot\|_{X_{\theta,q}} := \|\cdot\|_{X_0} + \rho(\cdot)_{\theta,q}$ is a quasi-norm for $X_{\theta,q}$. The reader can verify that also here, $X_{\theta,q}$ is a quasinormed linear complete space.

In both of these cases $X_{\theta,q}$ is a Banach space if X_0, X_1 have this property and $1 \le q \le \infty$.

If $t^{-\theta}K(f,t) \ge a > 0$ for some t, then $s^{-\theta}K(f,s) \ge 2^{-\theta}a$ for $t \le s \le 2t$. Hence, if f belongs to the space $(X_0, X_1)_{\theta,q}$, $0 < q < \infty$, then $t^{-\theta}K(f,t) \to 0$ for $t \to 0$ and for $t \to \infty$ because otherwise the integral in (7.2) would diverge. This together with (1.9) (or its analogue in the case K is defined by (1.10)) implies that when $\theta = 0, 1$ and $q < \infty$, $X_{\theta,q}$ consists only of functions in $\mathcal{N}$. We shall exclude these spaces in what follows.

In Chapter 2, we have had several *interpolation theorems*, Theorems 4.2, 4.4, 4.5. Here is one for the θ, q-spaces:

Theorem 7.1. *Let X_0, X_1 and Y_0, Y_1 be pairs of complete quasi-normed linear spaces. If U is a linear operator which maps X_i into Y_i with norm M_i, $i = 0, 1$, then U boundedly maps $X_{\theta,q}$ into $Y_{\theta,q}$ with norm $\le M_0^{1-\theta}M_1^{\theta}$, for all $0 < q \le \infty$ and $0 < \theta < 1$. If $|Uf|_{Y_1} \le M_1|f|_{X_1}$ for quasi-seminorms on X_1, Y_1, then U maps $X_{\theta,q}$ into $Y_{\theta,q}$ with norm not exceeding $\max(M_0, M_0^{1-\theta}M_1^{\theta})$.*

Proof. If $f \in X_0 + X_1$, then from (1.14)

$$K(Uf, t; Y_0, Y_1) \le M_0 K(f, t, M_1 t/M_0; X_0, X_1)\,.$$

If we now apply a θ, q norm to both sides of this inequality, we obtain $\rho(Uf)_{\theta,q} \le M_0^{1-\theta}M_1^{\theta}\rho(f)_{\theta,q}$ and the first statement of the theorem follows. In the seminorm case, we add $\|f\|_{Y_0}$ and $\|f\|_{X_0}$ to $\rho(Uf)_{\theta,q}$ and $\rho(f)_{\theta,q}$, respectively, and use $\|Uf\|_{Y_0} \le M_0\|f\|_{X_0}$ to obtain the second statement of the theorem. □

If X_1 is continuously embedded in X_0, the integral in (7.2) can be taken over $[0, a]$ for any fixed constant $a > 0$:

$$\rho(f)_{\theta,q} \sim \begin{cases} \left(\int_0^a \left[t^{-\theta}K(f,t)\right]^q \dfrac{dt}{t}\right)^{1/q}, & 0 < q < \infty \\ \sup\limits_{0 \le t < a} t^{-\theta}K(f,t)\,, & q = \infty \end{cases} \tag{7.4}$$

with constants of equivalence depending only on θ and q. Indeed, in this case, for $t \ge a$, we have for any $g \in X_1$ by the definition of K,

$$K(f,t) \le \|f\|_{X_0} \le \|f - g\|_{X_0} + \|g\|_{X_0} \le C_1\left[\|f - g\|_{X_0} + a\|g\|_{X_1}\right]$$

where C_1 is a fixed constant. Thus $K(f,t) \le C_1 K(f,a)$, $t \ge a$. Applying this to the integral (7.2), we see that $\int_a^\infty \left[t^{-\theta}K(f,t)\right]^q \frac{dt}{t} \le CK(f,a)^q$ and this can be absorbed by the integral over $[0, a]$. Similarly for $q = \infty$. The same argument shows that (7.4) also holds when the K-functional is defined using a quasi-seminorm satisfying (7.3).

It is sometimes convenient to use equivalent norms obtained from (7.4) by discretizing the θ, q norms. We mention the following two equivalences:

$$\rho(f)_{\theta,q} \sim \begin{cases} \left(\sum\limits_{n=N}^{\infty} \left[n^{r\theta}K(f, n^{-r})\right]^q \dfrac{1}{n}\right)^{1/q}, & 0 < q < \infty \\ \sup\limits_{n \ge N} n^{r\theta}K(f, n^{-r})\,, & q = \infty\,, \end{cases} \tag{7.5}$$

$$\rho(f)_{\theta,q} \sim \begin{cases} \left(\sum\limits_{n=N}^{\infty} \left[2^{nr\theta}K(f, 2^{-nr})\right]^q\right)^{1/q}, & 0 < q < \infty \\ \sup\limits_{n \ge N} 2^{nr\theta}K(f, 2^{-nr})\,, & q = \infty\,. \end{cases} \tag{7.6}$$

Here N and $r > 0$ are arbitrary but fixed and the constants of equivalency are independent of f. These equivalencies follow simply from the monotonicity of $K(f, \cdot)$.

Before describing $X_{\theta,q}$ for specific pairs, we develop some general properties of these spaces. We have the following continuous embeddings:

$$X_{\alpha,q} \subset X_{\beta,r} \text{ if } \alpha > \beta \text{ or if } \alpha = \beta \text{ and } q \le r\,. \tag{7.7}$$

They are derived from (7.6) in exactly the same manner that (10.6) and (10.7) of Chapter 2 were established.

Embeddings are inherited by interpolation spaces: if $X_0 \subset Y_0$ and $X_1 \subset Y_1$ are continuous embeddings for the pairs X_0, X_1 and Y_0, Y_1 then $X_{\theta,q}$ is continuously embedded in $Y_{\theta,q}$ and

$$\|f\|_{Y_{\theta,q}} \le M\|f\|_{X_{\theta,q}}\,, \quad 0 \le \theta \le 1\,; \quad 0 < q \le \infty\,. \tag{7.8}$$

Indeed, for some $M > 0$, $\|f\|_{Y_i} \le M\|f\|_{X_i}$, $i = 0, 1$. From the definition of the K-functional,

$$K(f, t; Y_0, Y_1) \le M\,K(f, t; X_0, X_1)\,, \quad t \ge 0\,,$$

and (7.8) follows by taking a (θ, q) norm. The inequality (7.8) also holds if the K-functional is defined using quasi-seminorms.

If we begin with a pair X_0, X_1 of quasi-normed spaces and generate the interpolation spaces $X_{\theta,q}$, we could apply interpolation again to a pair of the $X_{\theta,q}$. The question arises whether we obtain anything new in this way. The important *reiteration theorem* of interpolation says that we do not. Let $Y_0 := X_{\alpha_0,q_0}$ and $Y_1 := X_{\alpha_1,q_1}$ with $0 < \alpha_0 < \alpha_1 < 1$ and $0 < q_0, q_1 \le \infty$. We shall compare the K-functionals $K(f,t) := K(f,t;X_0,X_1)$ and $\widetilde{K}(f,t) := K(f,t;Y_0,Y_1)$. In the case that a quasi-seminorm $|\cdot|_{X_1}$ is used in the definition of K, we use $\rho(f)_{\alpha_1,q_1}$ in the definition of $\widetilde{K}$.

Theorem 7.2 (Holmstedt [1970]). *For the K-functionals in question, we have for any $f \in Y_0 + Y_1$ and $\delta := \alpha_1 - \alpha_0$*

$$\widetilde{K}(f,t^\delta) \sim \left(\int_0^t [s^{-\alpha_0}K(f,s)]^{q_0}\frac{ds}{s}\right)^{1/q_0} + t^\delta\left(\int_t^\infty [s^{-\alpha_1}K(f,s)]^{q_1}\frac{ds}{s}\right)^{1/q_1}. \tag{7.9}$$

with constants of equivalency independent of f and t and with the usual change from an integral to a supremum when one or both of q_0, q_1 are infinite.

Proof. We shall assume that $\|\cdot\|_{X_i}$, $i = 0,1$, are quasi-norms and that $q_0, q_1 < \infty$. The case when one of these is a quasi-seminorm or when q_0 or $q_1 = \infty$ requires very slight modifications which we leave to the reader. We fix t and first prove that the left side of (7.9) is less than a multiple of the right. Since $K(f,t)$ is not zero, we can select $f_i \in X_i$, $i = 0,1$ such that $f = f_0 + f_1$ and $\|f_0\|_{X_0} + t\|f_1\|_{X_1} \le 2K(f,t)$. Then for $s, t > 0$,

$$K(f_0,s) \le \|f_0\|_{X_0} \le 2K(f,t), \tag{7.10}$$

$$K(f_1,s) \le s\|f_1\|_{X_1} \le 2\frac{s}{t}K(f,t). \tag{7.11}$$

Since $f_0 = f - f_1$, (1.12) and (7.11) yield

$$K(f_0,s) \le C[K(f,s) + K(f_1,s)] \le C\,K(f,s) + C\frac{s}{t}K(f,t). \tag{7.12}$$

We can now estimate $\|f_0\|_{Y_0}$; from (7.2), $\|f_0\|_Y = \int_0^t + \int_t^\infty$. Using (7.12), we see that the first integral does not exceed a constant multiple of

$$\left(\int_0^t [s^{-\alpha_0}K(f,s)]^{q_0}\frac{ds}{s}\right)^{1/q_0} + \left(\int_0^t \left[s^{-\alpha_0}\frac{s}{t}K(f,t)\right]^{q_0}\frac{ds}{s}\right)^{1/q_0}. \tag{7.13}$$

The second term in (7.13) is $\le C\,t^{-\alpha_0}K(f,t)$ and therefore it can be incorporated into the first term. Similarly, from (7.10), the second integral $\int_t^\infty$ in the formula for $\|f_0\|_{Y_0}$ is $\le C\,t^{-\alpha_0}K(f,t)$, so that it can likewise be incorporated into the first term of (7.13). Using all this, we find that

$$\|f_0\|_{Y_0} \le C\left(\int_0^t [s^{-\alpha_0}K(f,s)]^{q_0}\frac{ds}{s}\right)^{1/q_0}.$$

A similar argument shows that

$$\|f_1\|_{Y_1} \le \left(\int_t^\infty [s^{-\alpha_1}K(f,s)]^{q_1}\frac{ds}{s}\right)^{1/q_1}.$$

Since $\widetilde{K}(f,t^\delta) \le \|f_0\|_{Y_0} + t^\delta\|f_1\|_{Y_1}$, we have proved that the left side of (7.9) is less than a multiple of the right.

For the other direction, we suppose that $f = f_0 + f_1$ with $f_i \in Y_i$, $i = 0,1$. Then $K(f,s) \le C[K(f_0,s) + K(f_1,s)]$, $s \ge 0$. From the embedding $Y_1 \subset X_{\alpha_1,\infty}$ of (7.7), we have $K(f_1,s) \le C\,s^{\alpha_1}\|f_1\|_{Y_1}$. Hence the first term on the right side of (7.9) does not exceed a multiple of

$$\left(\int_0^t [s^{-\alpha_0}K(f_0,s)]^{q_0}\frac{ds}{s}\right)^{1/q_0} + \|f_1\|_{Y_1}\left(\int_0^t [s^{\alpha_1-\alpha_0}]^{q_0}\frac{ds}{s}\right)^{1/q_0} \le \|f_0\|_{Y_0} + C\,t^\delta\|f_1\|_{Y_1}.$$

Using the inequality $K(f_0,s) \le C\,s^{\alpha_0}\|f_0\|_{Y_0}$, we obtain the same estimate for the second term on the right side of (7.9). Taking an infimum over all such decompositions $f = f_0 + f_1$, we find that the right side of (7.9) does not exceed a multiple of the left. □

As a corollary, we have

Theorem 7.3 (Peetre [B-1963]). *Let X_0, X_1 be a pair of spaces and $Y_0 := X_{\alpha_0,q_0}$ and $Y_1 := X_{\alpha_1,q_1}$ with $0 < \alpha_0 < \alpha_1 < 1$ and $0 < q_0, q_1 \le \infty$. Then, for any $0 < \theta < 1$, $0 < r \le \infty$, we have*

$$(Y_0,Y_1)_{\theta,r} = (X_0,X_1)_{\theta',r}, \quad \theta' := (1-\theta)\alpha_0 + \theta\alpha_1. \tag{7.14}$$

with equivalent topologies.

Proof. Let $\widetilde{\rho}(f)_{\theta,r}$ denote the θ,r norm applied to $\widetilde{K}(f,\cdot)$. We need to show that $\widetilde{\rho}_{\theta,r}(f) \sim \rho_{\theta',r}(f)$. We shall do this in the case $r < \infty$. When $r = \infty$, a simple change from integrals to supremums in the following proof applies.

Let $\overline{K}$ denote the K-functional for the pair $X_{\alpha_0,\infty}, X_{\alpha_1,q_1}$. Since $X_{\alpha_0,q}$ is continuously embedded in $X_{\alpha_0,\infty}$, we have $\overline{K}(f,\cdot) \le \widetilde{K}(f,\cdot)$. We apply (7.9) for $\overline{K}$. The first term on the right side of (7.9) is then a supremum norm; we get

$$t^{-\alpha_0}K(f,t) \le C\overline{K}(f,t^\delta) \le C\widetilde{K}(f,t^\delta).$$

Therefore, an application of the θ', r norm to K gives

$$\int_0^\infty \left[t^{-\theta'}K(f,t)\right]^r\frac{dt}{t} \le C\int_0^\infty \left[t^{\alpha_0-\theta'}\widetilde{K}(f,t^\delta)\right]^r\frac{dt}{t} = C\delta^{-1}\int_0^\infty \left[t^{-\theta}\widetilde{K}(f,t)\right]^r\frac{dt}{t},$$

where the last equality follows by a change of variables. This gives $\rho(f)_{\theta',r} \le C\widetilde{\rho}(f)_{\theta,r}$.

For the converse direction, we use (7.9). A change of variables gives

$$\begin{aligned}\widetilde{\rho}_{\theta,r}(f)^r &= \delta\int_0^\infty \left[t^{-\delta\theta}\widetilde{K}(f,t^\delta)\right]^r \frac{dt}{t}\\ &\le C\int_0^\infty \left(t^{-\theta\delta q_0}\int_0^t \left(s^{-\alpha_0}K(f,s)\right)^{q_0}\frac{ds}{s}\right)^{r/q_0}\frac{dt}{t}\\ &\quad + C\int_0^\infty \left(t^{(1-\theta)\delta q_1}\int_t^\infty \left(s^{-\alpha_1}K(f,s)\right)^{q_1}\frac{ds}{s}\right)^{r/q_1}\frac{dt}{t}\\ &=: CI_1 + CI_2 .\end{aligned}$$

To estimate I_1, we apply Hardy's inequality (3.1) of Chapter 2 (or in the case $r/q_0 < 1$, Theorem 3.5 of Chapter 2) with $q = r/q_0$ and $\phi(s) = [s^{-\alpha_0}K(f,s)]^{q_0}$ and obtain

$$I_1 \le \int_0^\infty \left[t^{-\theta\delta q_0}t^{-\alpha_0 q_0}K(f,t)^{q_0}\right]^{r/q_0}\frac{dt}{t} = \rho(f)^r_{\theta',r}$$

because $\theta\delta + \alpha_0 = \theta'$. A similar inequality holds for I_2 and hence $\widetilde{\rho}(f)_{\theta,r} \le C\,\rho(f)_{\theta',r}$. □

We give some examples of the interpolation spaces $X_{\theta,q}$.

1. *Interpolation spaces for* (L_1, L_∞) The K-functional for this pair is $\int_0^t f^*(s)ds$. Hence, if $0<\theta<1$, $0<q\le\infty$, from (3.7) of Chapter 2, we have $X_{\theta,q} = L_{p,q}$, where $\theta := 1 - 1/p$ and the following are equivalent norms:

$$\|f\|_{X_{\theta,q}} \sim \|f\|_{L_{p,q}} . \tag{7.15}$$

2. *Interpolation spaces for* (L_r, L_s), $r,s\ge 1$ From **1.** and the reiteration theorem, these are again the Lorentz spaces. Namely, $(L_r,L_s)_{\theta,q} = L_{p,q}$ for $1/p := (1-\theta)/r + \theta/s$ and $0<q\le\infty$.

3. *Interpolation spaces for* $(L_p(A), W^r_p(A))$, $A = \mathbb{R}$, $\mathbb{R}_+$, $\mathbb{T}$, *or* $[a,b]$ For $1\le p\le\infty$, $0<q\le\infty$, and the pair (L_p, W_p^r), the $\rho(\cdot)_{\theta,q}$ norm of (7.1) is equivalent to

$$\left(\int_0^\infty \left[t^{-\theta}\omega_r(f,t^{1/r})_p\right]^q \frac{dt}{t}\right)^{1/q} \sim |f|_{B_q^{\theta r}(L_p)}$$

where $B_q^\alpha(L_p)$ are the Besov spaces introduced in §10 of Chapter 2. Hence, $(L_p, W_p^r)_{\theta,q} = B_q^\alpha(L_p)$, $\alpha := \theta r$, $0<q\le\infty$.

4. We prove next a result of Peetre [1970] about powers of spaces which have been discussed in Proposition 1.3.

Proposition 7.4. *Let* (X_0,X_1) *be a pair of quasi-normed (or quasi-seminormed) spaces and let* $Y_i := X_i^{q_i}$, $i=0,1$, $0<q_i<\infty$. *If* $0<\theta<1$ *and*

$$\frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}, \quad \eta := \frac{\theta q}{q_1} = \frac{q-q_0}{q_1-q_0}, \tag{7.16}$$

then q *is between* q_0 *and* q_1 *and therefore* $0<\eta<1$. *Under these conditions the spaces*

$$(X_0,X_1)^q_{\theta,q} \text{ and } (Y_0,Y_1)_{\eta,1} \tag{7.17}$$

are identical and have equivalent power quasi-norms (or quasi-seminorms).

Proof. We use the notation of Proposition 1.3. According to the remark at the beginning of the present section, $f\in(X_0,X_1)^q_{\theta,q}$ implies $t^{-\theta}K(f,t)\to 0$ for $t\to 0$ and $t\to\infty$ and $f\in(Y_0,Y_1)_{\eta,1}$ implies $s^{-\eta}\overline{K}(f,s)\to 0$, $s\to 0$ or $s\to\infty$. This implies that the integration by parts which appear below have zero boundary terms.

Let f belong to either of the spaces (7.17). By means of partial integration, the substitution $s = t^{q_1}K(f,t)^{q_0-q_1}$, and (1.19) (with all integrals over $[0,\infty)$), we can now justify the following equivalences

$$\begin{aligned}\int s^{-\eta}\overline{K}(f,s)\frac{ds}{s} &\sim \int \overline{K}(f,s)d(s^{-\eta}) \sim \int K(f,t)^{q_0}d\left(t^{-\eta q_1}K(f,t)^{\eta(q_1-q_0)}\right)\\ &\sim \int t^{-\eta q_1}K(f,t)^{\eta(q_1-q_0)}d\left(K(f,t)^{q_0}\right)\\ &\sim \int t^{-\eta q_1}K(f,t)^{\eta(q_1-q_0)+(q_0-q)}d\left(K(f,t)^{q}\right)\\ &\sim \int t^{-\theta q}d\left(K(f,t)^q\right) \sim \int t^{-\theta q}K(f,t)^q\frac{dt}{t}.\end{aligned}$$

They show that the norms for the spaces (7.17) are equivalent. □

5. Finally, we consider the following sequence spaces which sometimes arise in the characterization of approximation classes. Let X be a complete quasi-normed linear space. If $\alpha>0$, $0<q\le\infty$, the space $l_q^\alpha(X)$ consists of all sequences $\boldsymbol{a} := (a_n)_0^\infty$, $a_n\in X$, $n=0,1,\dots$, such that the following is finite:

$$\|\boldsymbol{a}\|_{l_q^\alpha(X)} := \left(\sum_{n=0}^\infty [2^{n\alpha}\|a_n\|_X]^q\right)^{1/q}, \quad 0<q<\infty, \tag{7.18}$$

(with the obvious supremum in place of the sum if $q=\infty$). The reader can verify that each $l_q^\alpha(X)$ is a complete quasi-normed linear space. The following theorem summarizes the interpolation properties of these spaces.

Theorem 7.5. *Let* $0<q_0$, $q_1\le\infty$, $\alpha_0,\alpha_1>0$ *and let* $0<\theta<1$.

(i) *For* $\alpha := (1-\theta)\alpha_0 + \theta\alpha_1$, *and any* $0<q\le\infty$, *we have*

$$\left(l_{q_0}^{\alpha_0}(X),\ l_{q_1}^{\alpha_1}(X)\right)_{\theta,q} = l_q^\alpha(X). \tag{7.19}$$

(ii) *If X_0, X_1 are a pair of quasinormed spaces, and if*

$$\alpha := (1-\theta)\alpha_0 + \theta\alpha_1\,, \qquad \frac{1}{q} = \frac{1-\theta}{q_0} + \frac{\theta}{q_1}\,, \tag{7.20}$$

then

$$\left(l_{q_0}^{\alpha_0}(X_0), l_{q_1}^{\alpha_1}(X_1)\right)_{\theta,q} = l_q^{\alpha}\left((X_0,X_1)_{\theta,q}\right). \tag{7.21}$$

Proof. (i) We can assume that $\alpha_0 < \alpha_1$. We consider first the special case $q_0 = q_1 = \infty$. For each $n = 0, 1, \ldots$, if $a_n = a_n^0 + a_n^1$ with $a_n^i \in X$, $i = 0, 1$, then $\|a_n\|_X \le C(\|a_n^0\|_X + \|a_n^1\|_X)$ with $C = 1$ if $\|\cdot\|_X$ is a norm. Therefore, in computing the following infimum, we obtain, within a fixed multiplicative constant, the smallest value when a_n^0 and a_n^1 are multiples of a_n:

$$\inf_{a_n = a_n^0 + a_n^1} \left\{2^{n\alpha_0}\|a_n^0\|_X + t2^{n\alpha_1}\|a_n^1\|_X\right\} \sim \min\left(2^{n\alpha_0}, t2^{n\alpha_1}\right)\|a_n\|_X$$

with absolute constants of equivalency. This gives that

$$K(t) := K\left(\boldsymbol{a}, t; l_\infty^{\alpha_0}(X),\ l_\infty^{\alpha_1}(X)\right) \sim \sup_{n\ge 0} \min(2^{n\alpha_0}, t2^{n\alpha_1})\|a_n\|_X\,. \tag{7.22}$$

Since an l_∞ norm does not exceed an l_q norm, the q-th power of the right side of (7.22) does not exceed $\sum_{n=0}^\infty [\min(2^{n\alpha_0}, t2^{n\alpha_q})\|a_n\|_X]^q$. This gives

$$\begin{aligned} \int_0^\infty \left[t^{-\theta}K(t)\right]^q \frac{dt}{t} &\le C^q \sum_{n=0}^\infty \|a_n\|_X^q \int_0^\infty t^{-\theta q-1}\left[\min(2^{n\alpha_0}, 2^{n\alpha_1}t)\right]^q dt \\ &\le C^q \sum_{n=0}^\infty 2^{n\alpha q}\|a_n\|_X^q = C^q \|\boldsymbol{a}\|_{l_q^\alpha(X)}^q \end{aligned} \tag{7.23}$$

where $\alpha := (1-\theta)\alpha_0 + \theta\alpha_1$.

We can reverse this last inequality. Namely, for $t \in [2^{-n(\alpha_1-\alpha_0)}, 2^{-(n+1)(\alpha_1-\alpha_0)}]$, we have from (7.22) that $\|a_n\|_X 2^{n\alpha_0} \le CK(t)$. Therefore,

$$\|a_n\|_X^q 2^{nq\alpha_0} \int_{2^{-n(\alpha_1-\alpha_0)}}^{2^{-(n+1)(\alpha_1-\alpha_0)}} t^{-\theta q-1}dt \le C^q \int_{2^{-n(\alpha_1-\alpha_0)}}^{2^{-(n+1)(\alpha_1-\alpha_0)}} t^{-\theta q-1}K(t)^q dt\,. \tag{7.24}$$

The left side of (7.24) is $\ge C\|a_n\|_X^q 2^{nq\alpha}$. Therefore, summing the resulting inequalities for $n = 0, 1, \ldots$ gives the reverse inequality to (7.23) and proves (7.19) in our special case. The general case of (i) now follows by applying the reiteration Theorem 7.3.

(ii) We shall compute $(l_0, l_1)_{\eta,1}$ for $l_i := \left[l_{q_i}^{\alpha_i}(X_i)\right]^{q_i}$, $i = 0, 1$, and $\eta := \theta q/q_1$ which according to Proposition 7.4 is equal to $\left(l_{q_0}^{\alpha_0}(X_0),\ l_{q_1}^{\alpha_1}(X_1)\right)_{\theta,q}^q$. In computing $K(t) := K(\boldsymbol{a}, t; l_0, l_1)$ by considering all decomposition $\boldsymbol{a} = \boldsymbol{a}^0 + \boldsymbol{a}^1$, $\boldsymbol{a}_i \in l_i$, $i = 0, 1$, we can take the infimum over each coordinate separately. This gives

$$K(t) = \sum_{n=0}^\infty 2^{nq_0\alpha_0} K\left(a_n, 2^{n(\alpha_1 q_1 - \alpha_0 q_0)}t; X_0^{q_0}, X_1^{q_1}\right).$$

From this, and the identities $\eta q_1 = \theta q$, $(1-\eta)q_0 = (1-\theta)q$, we obtain by integration,

$$\begin{aligned} \int_0^\infty t^{-\eta}K(t)\frac{dt}{t} &= \sum_{n=0}^\infty 2^{n\alpha_0 q_0} 2^{n\eta(\alpha_1 q_1 - \alpha_0 q_0)} \int_0^\infty t^{-\eta} K\left(a_n, t; X_0^{q_0}, X_1^{q_1}\right) \frac{dt}{t} \\ &= \sum_{n=0}^\infty 2^{n\alpha q}\|a_n\|_Y \end{aligned} \tag{7.25}$$

where $Y := (Y_0, Y_1)_{\eta,1}$. By Proposition 7.4, $\|a_n\|_Y \sim \|a_n\|_{(X_0,X_1)_{\theta,q}}^q$ with constants of equivalency independent of n. Therefore, (7.25) gives $(l_0, l_1)_{\eta,1} = \left[l_q^\alpha\left((X_0,X_1)_{\theta,q}\right)\right]^q$. □

§8. Problems

8.1. For $0 < \alpha < 1$, we have $K(f, t; C, \mathrm{Lip}\,\alpha) = \frac{1}{2}\overline{\omega}(f, 2t)$, $t > 0$, where $\overline{\omega}(f,t)$ is the concave majorant of $\omega(f, t^{1/\alpha})$.

8.2. (Ditzian-Totik [A-1987]). If $f(x) := (1+x)^a|\log(1+x)|^b$ and $a > -1/p$, then for $\phi(x) = \sqrt{1-x^2}$,

$$\omega_r^\phi(f,t)_p \sim \begin{cases} t^{2a+2/p}(\log 1/t)^b\,, & \text{if } a \ne 0, 1, \ldots \\ t^{2a+2/p}(\log 1/t)^{b-1}\,, & \text{if } a = 0, 1, \ldots, b \ne 0\,. \end{cases}$$

8.3. If $0 < q \le \infty$, $1 \le p \le \infty$, and $\beta = \alpha + k$ with k a positive integer and $\alpha > 0$, then $f \in B_q^\beta(L_p)$ if and only if $f^{(k-1)}$ is absolutely continuous and $f^{(k)} \in B_q^\alpha(L_p)$.

8.4. If $1 \le p \le \infty$, $r = 1, 2 \ldots$, and $f \in W_p^r(I)$, $I := [0,1]$, then there is a function $f_1 \in W_p(\mathbb{R})$ with $f_1 = f$ on I and $\|f_1\|_{W_p^r(\mathbb{R})} \le C\|f\|_{W_p^r(I)}$ and C depending only on r.

8.5. For $A = [a,b]$ and Lebesgue measure, $K(f, t; L_1(A), C(A)) = K(f, t; L_1(A), L_\infty(A))$.

8.6. (DeVore-Scherer [1979]). For the Sobolev spaces $W_1^1(A)$, $W_\infty^1(A)$, $A = [a,b]$, we have $K\left(f, t; W_1^1, W_\infty^1\right) \sim \int_0^t [f^*(s) + f'^*(s)]\,ds$.

8.7. ([Mitjagin-Semenov [1977]) There is a kernel $K(x,y)$ such that the operator $L(f,x) := \int_{\mathbb{T}} f(y)K(x,y)dy$ is bounded on $C(\mathbb{T})$ and $W_\infty^2(\mathbb{T})$ but is not bounded on $\mathrm{Lip}(1, C(\mathbb{T}))$. Hence, $\mathrm{Lip}(1, C(\mathbb{T}))$ is not an interpolation space for the pair $C(\mathbb{T}), W_\infty^2(\mathbb{T})$.

§ 9. Notes

9.1. The theory of K-functionals and of the related interpolation can be extended to more general situations than those given in §1 and §7. Peetre and Sparr [1972] have done this in the context of quasi-normed abelian groups X. In this case the topology on X is given by a quasi-norm or the power of a quasi-norm (see [Peetre-Sparr, 1972] for details).

9.2. It is also possible to give a formula for the K-functional between some Sobolev spaces of the same smoothness. If $r = 1, 2, \dots$, then DeVore and Scherer [1979] have shown that on $A = [a, b]$:

$$K\left(f, t; W_1^r(A), W_\infty^r(A)\right) \sim \int_0^t \sum_{k=0}^r \left[f^{(k)}\right]^* (s) ds .$$

From this it follows that $(W_1^r, W_\infty^r)_{\theta,p} = W_p^r$ if $\theta = 1 - 1/p$ and $1 < p < \infty$. The interpolation for other pairs of Sobolev spaces of the same smoothness order can then be determined by the reiteration theorem.

9.3. There is another important extension theorem of Whitney. If A is a closed compact subset of $\mathbb{R}$, then f is in $\mathrm{Lip}(\alpha, A)$, $0 < \alpha \le 1$, provided $|f(x) - f(y)| \le M|x - y|^\alpha$, for all $x, y \in A$. Whitney [1957] has shown that it is always possible to extend such a function to all of $\mathbb{R}$ with the preservation of the Lipschitz condition. Whitney's theorem can also be generalized to $\alpha > 1$ and to compact subsets of $\mathbb{R}^n$ (see the book of Stein [B-1970, Chapter 6]). These generalizations have been used by Jonnson and Wallin [B-1984] in the study of smoothness spaces and the approximation by polynomials on general sets A.

9.4. Weighted moduli of smoothness are extensively studied in the monograph of Ditzian and Totik [A-1987]. They introduce such moduli for a general class of weights ϕ, develop their important properties and show their application to approximation, in particular to the characterization of functions with a prescribed order of approximation by algebraic polynomials. Other authors studied similar moduli. Most notably, K. Ivanov (see for example Ivanov[1980]) has introduced averaged weighted moduli (which are equivalent to the Ditzian-Totik moduli) and applied them to polynomial approximation.

9.5. The K-functionals give only one of the methods for generating interpolation spaces for a given pair X_0, X_1; there are others. The most notable of these is the complex method of interpolation developed by Calderón [1964]. The question arises whether it is possible to characterize all the interpolation spaces for X_0, X_1 by one of these methods. For the K functional, this is possible provided the pair X_0, X_1 satisfies the two additional properties that follow.

The space X_0 is said to be Gagliardo complete (in $X_0 + X_1$) if it contains the limits (in $X_0 + X_1$) of each sequence (g_n) such that $\sup \|g_n\|_{X_0} < \infty$; similarly for X_1. A normed linear space X for which $X_0 \cap X_1 \subset X \subset X_0 + X_1$ is said to be K-monotone if $K(f, t) \le K(g, t)$, $t > 0$, and $g \in X$ implies $f \in X$ and $\|f\|_X \le \|g\|_X$. Brudnyi and Kruglyak [1981] have shown that if X_0, X_1 are Gagliardo complete and if each interpolation space X of X_0, X_1 is K-monotone then the norm on X is equivalent to $\Phi(K(f, \cdot))$ where Φ is one of the admissible function norms (3.4) of Chapter 2.

One way of proving that all interpolation spaces X of a pair X_0, X_1 are K monotone is to show that for each $f, g \in X_0 + X_1$ which satisfy $K(f, t) \le K(g, t)$, $t > 0$, there is a bounded linear operator U on X_0 and X_1 which maps g into $f : U(g) = f$. This has been done for certain pairs of spaces; most notably pairs of L_p spaces (Calderón [1966], Lorentz and Shimogaki [1971]). On the other hand, it is known that L_p, W_p^1, $1 < p \le \infty$ do not have the K-monotonicity property (Sedaev and Semenov [1971]). See Cwikel [1976] and the references therein for details.

Chapter 7. Central Theorems of Approximation

§ 1. Introduction

This chapter could be called the core of this book. Indeed, it establishes first theorems on the rate of approximation, for given real functions, by the basic approximation classes: by trigonometric polynomials, by algebraic polynomials, by splines. In many cases, finer results are given in subsequent chapters. Very often, the approximants can be realized through a sequence of linear operators. Usually, it is impossible in this way to obtain best possible approximation (see Theorem 4.1 of Chapter 9), so we must be prepared to accept just good approximation.

To explain the two main types of approximation theorems, we quote the theorem of Jackson (1912) concerning uniform trigonometric approximation on $\mathbb{T}$,

$$\text{(1.1)} \quad \text{If } f \in C^r(\mathbb{T}), \text{ then} \quad E_n(f) \le C_r n^{-r} \omega\left(f^{(r)}, n^{-1}\right), \qquad n = 1, 2, \dots,$$

and that of Bernstein,

$$\text{(1.2)} \quad \begin{gathered} \text{if for some } 0 < \alpha < 1, \quad E_n(f) \le C_r n^{-r-\alpha}, \quad n = 1, 2, \dots, \\ \text{then } f^{(r)} \in \mathrm{Lip}\,\alpha . \end{gathered}$$

Here, (1.1) is a *direct approximation theorem*, which asserts that smoothness of the function f implies a quick decrease to zero of its error of approximation by trigonometric polynomials. The *inverse theorem* (1.2) has the opposite implication. In ideal cases, the two theorems match each other. For example, it follows from (1.1) and (1.2) that $E_n(f) = \mathcal{O}(n^{-\alpha})$, $0 < \alpha < 1$ is equivalent to $f \in \mathrm{Lip}\,\alpha$.

Jackson theorems are proved in §2 by means of special integral operators. But there is another more sophisticated approach to direct theorems, §4, due to Favard. Its advantage is that it gives better, sometimes the best, constants in direct estimates.

From the trigonometric, we obtain in §6 direct theorems of algebraic polynomial approximation on an interval $[a, b]$, which may be assumed to be $[-1, 1]$. Inverse theorems are conspicuously absent in this section. The reason for this is that the estimates of §6 are not best possible. There exist finer theorems, with improved polynomial approximation near the endpoints ± 1; this finer theory will be treated in Chapter 8.

The bounds for the approximation error in the direct theorems for functions $f \in L_p$ will usually be of one of the forms

$$(1.3) \qquad E_n(f)_p \le \begin{cases} C_r n^{-r} \|f^{(r)}\|_p & (A) \\ C_{\alpha,p} n^{-\alpha} \|f\|_{\mathrm{Lip}^*(\alpha,p)} & (B) \\ C_r n^{-k}\, \omega_{r-k}(f^{(k)}, n^{-1})_p, \quad 0 \le k \le r & (C) \\ C_r \omega_r(f, n^{-1})_p & (D). \end{cases}$$

For analytic functions, the error estimates of §8 are much smaller. If one is not concerned about constants, the implications $(D) \to (C) \to (B) \to (A)$ follow from elementary properties of moduli of smoothness of Chapter 2. On the other hand, the theorems of §5 show that one has $(A) \to (D)$ if no side conditions are present. In this sense, the four types of inequalities are equivalent and it is enough to establish the simplest of these (A).

An abstract approach to estimates of the form (1.3) is given in §5. This section can be read on two levels. For linear normed spaces X, its results will be used in this and the next chapter. For quasi-normed spaces X, they are important for the last two sections of Chapter 12.

Sections 2–4 give a relatively complete presentation of the fundamental results of approximation by trigonometric polynomials (elementary in the sense that §5 is not used). In §8 we do the same for analytic functions defined on a real interval, the approximation tool being, naturally enough, algebraic polynomials. The abstract theorems of §5 allow one to upscale the conclusion of theorems according to the table (1.3). This will be often used in this book. Sections 6–7 for the polynomial and for spline approximation serve only as introductions to their subjects. They should be followed by the examination of Chapter 8 and of Chapter 12, respectively.

In §9, we define spaces which depend upon the asymptotic behavior of the approximation errors $E_n(f)$. Interesting in themselves, they will play an important role in the second half of Chapter 12.

§ 2. Trigonometric Approximation

We discuss here approximation to $f \in L_p(\mathbb{T})$, $1 \le p \le \infty$, from the space $\mathcal{T}_n$ of trigonometric polynomials of degree $\le n$ and its error $E_n(f) := E_n(f)_p$. When $p = \infty$, $L_\infty(\mathbb{T})$ is replaced by $C(\mathbb{T})$.

With Jackson (in 1912, see Jackson [A-1930]) we use the integrals of the following type

$$(2.1) \qquad L_n(f,x) := \int_{\mathbb{T}} f(x+t) \Lambda_n(t)\, dt = \int_{\mathbb{T}} f(t) \Lambda_n(x-t)\, dt$$

where Λ_n is an *even, non-negative* trigonometric polynomial,

$$\Lambda_n(t) = \sum_0^n a_k \cos kt \ge 0\,, \qquad t \in \mathbb{T}\,,$$

of degree $\le n$ with $\int_{-\pi}^{\pi} \Lambda_n\, dt = 1$. Writing $\cos k(x-t) = \cos kx \cos kt + \sin kx \sin kt$ and using the last integral in (2.1) we see that $L_n f$ is a trigonometric polynomial of degree $\le n$, which is even if f is even.

Since Λ_n is even,

$$\begin{aligned} L_n(f,x) - f(x) &= \int_{\mathbb{T}} [f(x+t) - f(x)]\, \Lambda_n(t)\, dt \\ &= \int_0^\pi [f(x+t) + f(x-t) - 2f(x)]\, \Lambda_n(t)\, dt\,. \end{aligned}$$

If the kernel Λ_n satisfies the inequalities

$$(2.2) \qquad \int_0^\pi t^k \Lambda_n(t)\, dt \le C n^{-k}\,, \qquad k = 0,1,2\,,$$

we can use the relation $\omega_2(f,t)_p \le (nt+1)^2 \omega_2\left(f, \frac{1}{n}\right)_p$ (see (5.13) of Chapter 2) and find, in the $L_p(\mathbb{T})$ norm, by Minkowski's inequality (1.6) of Chapter 2,

$$(2.3) \qquad \begin{aligned} \|f - L_n f\| &\le \int_0^\pi \|f(x+t) + f(x-t) - 2f(x)\| \Lambda_n(t)\, dt \le \int_0^\pi \omega_2(f,t) \Lambda_n(t)\, dt \\ &\le \omega_2\left(f, n^{-1}\right) \int_0^\pi (nt+1)^2 \Lambda_n(t)\, dt \le C\omega_2\left(f, n^{-1}\right)\,. \end{aligned}$$

One of the simplest kernels of this type is the Jackson kernel

$$(2.4) \qquad K_n(t) := \lambda_n \left(\frac{\sin mt/2}{\sin t/2}\right)^4, \quad m := \left[\frac{n}{2}\right] + 1\,, \quad \int_{\mathbb{T}} K_n\, dt = 1\,;$$

the last equality defines the constant λ_n. The *Jackson integral* is given by

$$J_n(f,x) = \int_{\mathbb{T}} f(x+t) K_n(t)\, dt\,.$$

Also, important are the generalized Jackson kernels:

$$(2.5) \qquad K_{n,r}(t) := \lambda_{n,r} \left(\frac{\sin mt/2}{\sin t/2}\right)^{2r}, \quad \int_{\mathbb{T}} K_{n,r}\, dt = 1$$

where $n, r > 0$, $m := [n/r] + 1$, and $\lambda_{n,r}$ is defined by the last relation in (2.5). Since by (1.12) of Chapter 1,

$$\frac{1}{2m} \left(\frac{\sin mt/2}{\sin t/2}\right)^2 = \frac{1}{2} + \sum_0^{m-1} (1 - k/m) \cos kt\,,$$

it follows that $K_{n,r}$ is an even, non-negative trigonometric polynomial of degree $\le n$.

Lemma 2.1. *For each $r = 1, 2, \dots$, we have the equivalence $\lambda_{n,r} \sim n^{-2r+1}$ as $n \to \infty$; and there is a constant C_r for which*

$$\int_0^\pi t^k K_{n,r}(t)\,dt \le C_r n^{-k}, \qquad k = 0, 1, \dots, 2r-2. \tag{2.6}$$

Proof. Since $t/\pi \le \sin t/2 \le t/2$ for $0 \le t \le \pi$, we have

$$\lambda_{n,r}^{-1} = 2\int_0^\pi \left(\frac{\sin mt/2}{\sin t/2}\right)^{2r} dt \sim \int_0^\pi \left(\frac{\sin mt/2}{t}\right)^{2r} dt$$
$$\sim m^{2r-1} \int_0^{m\pi/2} \left(\frac{\sin u}{u}\right)^{2r} du \sim n^{2r-1}.$$

Hence $\lambda_{n,r} \sim n^{-2r+1}$. Similarly, for $0 < k \le 2r-2$,

$$\int_0^\pi t^k K_{n,r}(t)\,dt \le C\lambda_{n,r} \int_0^\pi t^k \left(\frac{\sin mt/2}{t}\right)^{2r} dt$$
$$\le C n^{2r-k-1}\lambda_{n,r} \int_0^{m\pi/2} u^k \left(\frac{\sin u}{u}\right)^{2r} du \le Cn^{-k}. \qquad \square$$

In particular, $\Lambda_n := K_{n,2}$ satisfies (2.2) with $k = 1, 2$, and we obtain from (2.3):

Theorem 2.2 (Jackson). *For some absolute constant C,*

$$\|f - J_n(f)\|_p \le C\omega_2\left(f, n^{-1}\right)_p \le C\omega\left(f, n^{-1}\right)_p. \tag{2.7}$$

To replace here ω_2 by $\omega_r, r > 2$ we use the operators:

$$S_n(f,x) := S_{n,r}(f,x) := \int_{\mathbb{T}} \left[(-1)^{r+1}\Delta_t^r(f,x) + f(x)\right] K_{n,r}(t)\,dt \tag{2.8}$$

where $\Delta_t^r(f,x) = \sum_0^r (-1)^{r+k}\binom{r}{k} f(x+kt)$ are the r^{th} differences of f. Note that $S_n(f)$ is a linear combination of terms

$$\int_{\mathbb{T}} f(x+kt)\cos lt\,dt \qquad k = 1, \dots, r;\ l = 1, \dots, n. \tag{2.9}$$

Since $f(x+kt)$ has period $2\pi/k$ as a function of t, according to Lemma 10.3 of Chapter 3, (2.9) is zero unless k divides l. In the latter case, substitution $u = x + kt$ yields that (2.9) is a trigonometric polynomial of degree l/k. It follows that $S_n f$ is a trigonometric polynomial of degree $\le n$.

Theorem 2.3 (Stechkin [1951]). *For $r = 1, 2, \dots$, there is a constant C_r such that*

$$E_n(f)_p \le C_r\omega_r\left(f, n^{-1}\right)_p, \qquad n = 1, 2, \dots, \quad 1 \le p \le \infty, \tag{2.10}$$

whenever $f \in L_p(\mathbb{T}), 1 \le p < \infty$; or $f \in C(\mathbb{T})$, $p = \infty$.

Proof. Using the inequalities $\omega_r(f,t) \le (nt+1)^r\omega_r(f,1/n)$ and (2.6), we find in the L_p norm, via Minkowski's inequality,

$$\begin{aligned} \|S_n(f) - f\| &\le \left\| \int_{\mathbb{T}} \Delta_t^r(f,x) K_{n,r}(t)\,dt \right\| \le \int_{\mathbb{T}} \omega_r(f,|t|) K_{n,r}(t)\,dt \\ &\le \omega_r(f,n^{-1}) \int_{\mathbb{T}} (n|t|+1)^r K_{n,r}(t)\,dt \le C\,\omega_r(f,n^{-1}). \end{aligned} \tag{2.11}$$ $\square$

Corollary 2.4. *If $f \in W_r^p$, then*

$$E_n(f)_p \le C_r\, n^{-r}\omega\left(f^{(r)}, 1/n\right)_p. \tag{2.12}$$

Corollary 2.5. *If $f \in \mathrm{Lip}^*(\alpha,p)$, $\alpha > 0$, then $E_n(f)_p = \mathcal{O}(n^{-\alpha})$.*

For $f \in \mathrm{Lip}^*(\alpha,p)$ is equivalent to $\omega_{[\alpha]+1}(f,h)_p \le Mh^\alpha$. $\square$

Remark. Theorems 2.2 and 2.3 and their consequences remain valid for functions of any rearrangement-invariant space X on $\mathbb{T}$ (see §2 of Chapter 2). We define the modulus of smoothness on X by

$$\omega_r(f,h)_X = \sup_{|t| \le h} \|\Delta_t^r f(\cdot)\|_X.$$

The translation operator $f(x) \to f(x+a)$ is norm-preserving on X; if it is continuous, then the inequality (7.8) of Chapter 2 is valid. We can repeat the calculations (2.11) and obtain

$$E_n(f)_X \le C\omega_r\left(f, n^{-1}\right)_X. \tag{2.13}$$

It is easy to see that the translation operator is continuous in each rearrangement invariant space X on $\mathbb{T}$ which has an *absolutely continuous norm*: $\|f\chi_B\|_X \to 0$ when $|B| \to 0$ for each $f \in X$.

The following results are for the spaces $X = L_p(\mathbb{T})$, $1 \le p < \infty$, and for $X = C(\mathbb{T})$ if $p = \infty$.

Trignometric polynomials $\widetilde{S}_n$ with mean value zero, that is, with the constant term $a_0 = 0$, possess an integral on $\mathbb{T}$, which also has mean value zero. If $f \in L_p$ has mean value zero (for instance, if it is a derivative), if $\widetilde{S}_n$ is an arbitrary trigonometric polynomial with the constant term a_0, then

$$|a_0| \le \frac{1}{2\pi}\int_{\mathbb{T}} |f(t) - \widetilde{S}_n(t)|\,dt \le (2\pi)^{-1/p}\|f - \widetilde{S}_n\|_p\,. \tag{2.14}$$

As a corollary,

1. *If f has mean value zero, and if S_n is derived from $\widetilde{S}_n$ by omitting its constant term, then*

$$\|f - S_n\|_p \le 2\|f - \widetilde{S}_n\|_p\,, \qquad 1 \le p \le \infty\,. \tag{2.15}$$

2. For $f \in W_p^1$ we have, with an absolute constant C,

$$E_n(f)_p \le C\frac{1}{n}E_n(f')_p\,, \qquad 1 \le p \le \infty\,. \tag{2.16}$$

For the proof, let $\widetilde{S}_n$ be the best approximation to f', let S'_n be $\widetilde{S}_n$ with the constant term removed, and let S_n be the periodic integral of S'_n. By (2.7) and **1** we have then

$$\begin{aligned} E_n(f)_p = E_n(f - S_n)_p \le C\omega_1(f - S_n, 1/n)_p \le C\frac{1}{n}\|f' - S'_n\|_p \\ \le 2C\frac{1}{n}\|f' - \widetilde{S}_n\|_p = 2C\frac{1}{n}\,E_n(f')_p\,, \end{aligned}$$

where C is the best constant of (2.7).

3. *By iteration we obtain, with C_r depending on r, Jackson's relation*

$$E_n(f)_p \le C_r n^{-r} E_n\left(f^{(r)}\right)_p\,, \qquad f \in W_p^r\,. \tag{2.17}$$

From this, Corollary 2.4 follows again.

Many authors have discussed *simultaneous approximation* of f and its derivatives $f^{(k)}$, $k = 1, \dots, r$, by T_n and its derivatives $T_n^{(k)}$. They follow easily from the above if one uses a lemma by Zamansky [1949]:

4. Lemma 2.6. *For each $g \in L_p$, $1 \le p < \infty$, or $g \in C$ for $p = \infty$, and each polynomial T_n that satisfies*

$$\|g - T_n\|_p \le K\omega(g, 1/n)_p \tag{2.18}$$

one has, for $K_1 = 2(K + 1)$,

$$\|T'_n\|_p \le K_1 n\omega(g, 1/n)_p\,. \tag{2.19}$$

Proof. Let $h := 1/n$. From (2.18), omitting subscripts p,

$$\|T_n(\cdot + h) - T_n(\cdot - h)\| \le 2\|g - T_n\| + \omega(g, 2h) \le (2K + 2)\omega(g, h)\,. \tag{2.20}$$

On the other hand, using the Taylor series for T_n and afterwards the Bernstein L_p-inequality, one obtains

$$\begin{aligned} \|T_n(\cdot + h) - T_n(\cdot - h)\| &= \left\| \sum_{j=1}^{\infty} 2T_n^{(2j-1)} h^{2j-1}/(2j-1)! \right\| \\ &\ge 2h\|T'_n\| - 2\sum_{j=2}^{\infty} \|T_n^{(2j-1)}\| h^{2j-1}/(2j-1)! \\ &\ge 2h\|T'_n\| \left(1 - \sum_{j=2}^{\infty} (hn)^{2j-2}/(2j-1)!\right) \\ &= 2h\|T'_n\| \left(1 - \sum_{j=2}^{\infty} 1/(2j-1)!\right) \ge h\|T'_n\|\,. \end{aligned} \tag{2.21}$$

From these two inequalities we derive (2.19). □

5. From this we obtain a theorem about simultaneous approximation (Czipszer and Freud [1958]):

Theorem 2.7. *For each $f \in W_p^r$, $1 \le p \le \infty$, its polynomial T_n of best approximation in the L_p norm satisfies*

$$\|f^{(k)} - T_n^{(k)}\|_p \le C_r\, E_n(f^{(k)})_p\,, \qquad k = 0, \dots, r\,,\ n = 0, 1, \dots\,. \tag{2.22}$$

Proof. This is trivial if $r = 0$. Suppose the theorem is true for some r; we show that it is true with r replaced by $r + 1$.

Let $f \in W_p^{r+1}$ and let $\widetilde{S}_n$ be the polynomial of best approximation to f', let S_n be the periodic integral of $\widetilde{S}_n - a_0$. By (2.15) and by the inductive hypothesis we have, omitting subscripts r, p,

$$\|f^{(k+1)} - S_n^{(k+1)}\| \le CE_n(f^{(k+1)})\,, \qquad k = 0, \dots, r\,. \tag{2.23}$$

Let R_n be the best approximation to $f - S_n$. Clearly, $T_n = R_n + S_n$. By Jackson's estimate (2.7), by Lemma 2.6, and by the case $k = 0$ of (2.23),

$$\|R'_n\| \le C\|f' - S'_n\| \le CE_n(f')\,.$$

Therefore by the Bernstein inequality and **3**,

$$\|R_n^{(k+1)}\| \le Cn^k E_n(f') \le CE_n(f^{(k+1)})\,, \qquad k = 0, \dots, r\,.$$

We have $\|f - T_n\| = E_n(f)$ and moreover for $k = 0, \dots, r$,

$$\|f^{(k+1)} - T_n^{(k+1)}\| \le \|f^{(k+1)} - S_n^{(k+1)}\| + \|R_n^{(k+1)}\| \le CE_n(f^{(k+1)})\,. \quad □$$

The disadvantage of Theorem 2.7 is that polynomials of best approximation are rarely known. But all reasonable approximations will do:

Theorem 2.8. *Let $f \in W_p^r$ and $S_n \in \mathcal{T}_n$ have the properties*

$$(2.24) \qquad \|f - S_n\|_p \le C_r n^{-r} \omega\left(f^{(r)}, n^{-1}\right)_p, \qquad n = 1, 2, \dots,$$

then also for $n = 1, 2, \dots$

$$(2.25) \qquad \|f^{(k)} - S_n^{(k)}\|_p \le C_r n^{k-r} \omega\left(f^{(r)}, n^{-1}\right)_p, \qquad k = 1, \dots, r,$$

and

$$(2.26) \qquad \|S_n^{(r+1)}\|_p \le C_r n \omega\left(f^{(r)}, n^{-1}\right).$$

Proof. The best approximants T_n of f also satisfy (2.24). Therefore

$$\|S_n - T_n\| \le 2C_r n^{-r} \omega\left(f^{(r)}, n^{-1}\right).$$

Then Bernstein's inequality for $S_n - T_n$ and the relations $\|f^{(k)} - T_n^{(k)}\| \le C_r n^{k-r} \omega\left(f^{(r)}, 1/n\right)$ lead to (2.25). Finally, (2.26) follows by applying Lemma 2.6 to the relation (2.25) with $k = r$. □

§ 3. Inverse Theorems of Trigonometric Approximation

They have been initiated by Bernstein. Using his method of proof and the L_p-Bernstein inequality of §2, Chapter 4, we shall estimate $\omega_r(f, t)_p$ in terms of the errors of approximation $E_k(f)_p$.

Theorem 3.1. *For $f \in L_p(\mathbb{T})$, $1 \le p \le \infty$, and $r = 1, 2, \dots$, one has, with a constant C_r,*

$$(3.1) \qquad \omega_r\left(f, n^{-1}\right)_p \le C_r n^{-r} \sum_{k=1}^{n} k^{r-1} E_k(f)_p, \qquad n = 1, 2, \dots.$$

Proof. Let T_n stand for the polynomials of best approximation to f. For any $m = 0, 1, \dots$, and $h := 1/n$,

$$(3.2) \qquad \omega_r(f, h) \le \omega_r\left(f - T_{2^{m+1}}, h\right) + \omega_r\left(T_{2^{m+1}}, h\right).$$

Here we have

$$(3.3) \qquad \omega_r\left(f - T_{2^{m+1}}, h\right) \le 2^r E_{2^{m+1}}(f)$$

and

$$\begin{aligned}
\omega_r\left(T_{2^{m+1}}, h\right) \le h^r \|T_{2^{m+1}}^{(r)}\| &\le h^r \left\{ \|T_1^{(r)} - T_0^{(r)}\| + \sum_{\nu=0}^{m} \|T_{2^{\nu+1}}^{(r)} - T_{2^\nu}^{(r)}\| \right\} \\
&\le h^r \left\{ 2E_0(f) + \sum_{\nu=0}^{m} 2^{(\nu+1)r} \|T_{2^{\nu+1}} - T_{2^\nu}\| \right\} \\
&\le 2h^r \left\{ E_0(f) + \sum_{\nu=0}^{m} 2^{(\nu+1)r} E_{2^\nu}(f) \right\}.
\end{aligned}$$

Since $2^{(\nu+1)r} E_{2^\nu}(f) \le 2^{2r} \sum_{k=2^{\nu-1}+1}^{2^\nu} k^{r-1} E_k(f)$, $\nu \ge 1$, we derive from this

$$(3.4) \qquad \omega_r\left(T_{2^{m+1}}, h\right) \le 2^{2r+1} h^r \left\{ E_0(f) + E_1(f) + \sum_{k=2}^{2^m} k^{r-1} E_k(f) \right\}.$$

For a given n, we select m so that $2^m \le n < 2^{m+1}$. Then (3.2)–(3.4) yield (3.1) with the additional term $E_0(f)$. However, this term can be dropped, if one writes the obtained result for $f - c$, where c is the best constant approximation to f. □

As a special case, $r = 1$, we have

$$(3.5) \qquad \omega\left(f, n^{-1}\right)_p \le C n^{-1} \sum_{k=1}^{n} E_k(f)_p, \qquad f \in L_p.$$

The following theorem gives sufficient conditions for f to belong to W_p^r:

Theorem 3.2. *If for a function $f \in L_p$, $1 \le p \le \infty$, $r = 1, 2, \dots$,*

$$(3.6) \qquad \sum_{k=1}^{\infty} k^{r-1} E_k(f)_p < \infty,$$

then $f \in W_p^r$.

Proof. As before, for the polynomials of best approximation, $\|T_{2^{\nu+1}} - T_{2^\nu}\| \le 2E_{2^\nu}(f)$, and

$$\|T_{2^{\nu+1}}^{(r)} - T_{2^\nu}^{(r)}\| \le 2^{(\nu+1)r+1} E_{2^\nu}(f) \le C \sum_{k=2^{\nu-1}}^{2^\nu} k^{r-1} E_k(f).$$

It follows that $\sum \|T_{2^{\nu+1}}^{(r)} - T_{2^\nu}^{(r)}\| < \infty$, so that T_{2^ν} is a Cauchy sequence in $W_p^r(\mathbb{T})$. Since $T_{2^\nu} \to f$ in L_p, $f \in W_p^r(\mathbb{T})$. □

As an application of Theorem 3.1 we obtain that

$$(3.7) \qquad E_n(f)_p = \mathcal{O}(n^{-\alpha}), \qquad \alpha > 0, \ n = 1, 2, \dots$$

implies that

$$\omega_r(f,t) = \begin{cases} \mathcal{O}(t^\alpha) & \text{if } \alpha < r \\ \mathcal{O}\left(t^\alpha \log(1/t)\right) & \text{if } \alpha = r \\ \mathcal{O}(t^r) & \text{if } \alpha > r\,. \end{cases} \tag{3.8}$$

The number $r := [\alpha] + 1$, used in the definition of the spaces $\mathrm{Lip}^*(\alpha,p) := \mathrm{Lip}^*(\alpha, L_p)$ (see §9 of Chapter 2) satisfies $r > \alpha$, hence (3.7) implies $f \in \mathrm{Lip}^*(\alpha,p)$. Combining this with Corollary 2.5, we get

Theorem 3.3. *For $\alpha > 0$ and $1 \le p \le \infty$, relation $E_n(f) = \mathcal{O}(n^{-\alpha})$ is equivalent to $f \in \mathrm{Lip}^*(\alpha,p)$.*

For example, $E_n(f) = \mathcal{O}(1/n)$ is equivalent to $f \in Z_{1,p} := \mathrm{Lip}^*(1,p)$; but it does not imply that f belongs to the smaller space $\mathrm{Lip}(1,p)$, which is W_p^1 if $1 < p \le \infty$.

Almost without change, the proof of Theorem 3.1 applies in more general situations, and gives then an estimate of the K-functional of f, see Theorem 5.1(ii). Another interesting fact is that (3.1) yields a new proof of the Marchaud inequality (8.1) of Chapter 2 on $\mathbb{T}$. Indeed, using (2.10) to replace $E_k(f)_p$ by $\omega_s\left(f, \frac{1}{k}\right)_p$, with $s > r$, in (3.1), we obtain

$$\omega_r\left(f, n^{-1}\right)_p \le \frac{C_s}{n^r} \sum_{k=1}^{n} k^{r-1} \omega_s\left(f, k^{-1}\right)_p\,,$$

which is equivalent to (8.1), Chapter 2.

However, inequality (3.1) is not the best possible for $1 < p < \infty$! According to M. Timan [1958] and Zygmund one has

Theorem 3.4 (Timan-Zygmund). *For $f \in L_p(\mathbb{T})$, $1 < p < \infty$,*

$$\omega_r\left(f, n^{-1}\right)_p \le \begin{cases} \dfrac{C_r}{n^r}\left(\displaystyle\sum_{k=1}^{n} k^{rp-1}\, E_k(f)_p^p\right)^{1/p}, & 1 < p \le 2 \\ \dfrac{C_r}{n^r}\left(\displaystyle\sum_{k=1}^{n} k^{2r-1}\, E_k(f)_p^2\right)^{1/2}, & 2 \le p < \infty\,. \end{cases} \tag{3.9}$$

Proof. We write the improved Marchaud inequality (8.14) of Chapter 2 for $A = \mathbb{T}$ in its discrete form

$$\omega_r(f,n^{-1})_p \le C_r n^{-r}\left(\sum_{k=1}^{n} k^{r\mu-1}\omega_s(f,k^{-1})_p^\mu\right)^{1/\mu}, \quad \mu := \min(p,2)\,. \tag{3.10}$$

We take $s > r\mu$. The coefficients of the inequality (see (3.1))

$$\omega_s(f,k^{-1})_p \le C_s k^{-s} \sum_{j=1}^{k} j^{s-1} E_j(f)_p \tag{3.11}$$

satisfy $\sum_{j=1}^{k} k^{-s} j^{s-1} = \mathcal{O}(1)$. Applying the convex function $\varphi(u) := u^\mu$ to both sides of (3.11) by Jensen's inequality we obtain

$$\omega_s(f,k^{-1})_p^\mu \le C_s k^{-s} \sum_{j=1}^{k} j^{s-1} E_j(f)_p^\mu\,.$$

Substituting this into (3.10) and changing the order of summation, we obtain

$$\begin{aligned} \omega_r(f,n^{-1})_p &\le C_r' n^{-r}\left\{\sum_{k=1}^{n} k^{r\mu-1-s} \sum_{j=1}^{k} j^{s-1} E_j(f)_p^\mu\right\}^{1/\mu} \\ &\le C_r' n^{-r}\left\{\sum_{j=1}^{n} j^{r\mu-1} E_j(f)_p^\mu\right\}^{1/\mu}. \end{aligned}$$

□

However, (3.9) is only a slight improvement compared to (3.1): For $E_n(f)_p = \mathcal{O}(n^{-\alpha})$, both give $\omega_r(f,t)_p = \mathcal{O}(t^\alpha)$, if $r > \alpha$. If $r = \alpha$, (3.9) gives $\omega_r(f,t)_p = \mathcal{O}(t^\alpha(\log(1/t))^\sigma)$, $\sigma = \max(1/p, 1/2)$, which is better than (3.8).

§4. Favard's Theorems

Favard's theorems give better estimates of the error of trigonometric approximation than those of §2 in that they provide much better constants in the upper bounds for the error $E_n(f)$. They are applicable, however, to only very special classes of functions on $\mathbb{T}$.

By B_p^r, $1 \le p \le \infty$, $r = 1,2,\ldots$, we denote the class of functions $f \in W_p^r(\mathbb{T})$ for which $\|f^{(r)}\|_p \le 1$. Our starting point is the formula (5.14) of Chapter 5,

$$f(x) = \frac{1}{\pi}\int_{\mathbb{T}} f^{(r)}(t)\,\mathcal{B}_r(x-t)\,dt \tag{4.1}$$

valid for $f \in W_p^r(\mathbb{T})$ with mean value zero, and the periodic Bernoulli spline $\mathcal{B}_r$. With the Favard numbers K_r of (5.5) of Chapter 5, we shall prove, for the error of approximation of the class B_p^r by trigonometric polynomials of degree $\le n-1$:

$$\sup_{f\in B_p^r} E_{n-1}(f) =: E_{n-1}(B_p^r) \le K_r\, n^{-r}\,; \tag{4.2}$$

for $p = 1$ and $p = \infty$, one has equality in (4.2). This has an equivalent formulation in terms of individual functions:

$$E_{n-1}(f)_p \le K_r\, n^{-r}\|f^{(r)}\|_p\,, \qquad f \in W_p^r\,. \tag{4.3}$$

Since the addition of a constant to f does not change $E_{n-1}(f)_p$, we shall restrict ourselves to functions f on $\mathbb{T}$ with mean value zero: $\int_{\mathbb{T}} f\,dt = 0$. Formula (4.1) establishes a 1-1 correspondence between functions $f \in B_p^r$ and $\phi := f^{(r)} \in L_p$, $\|\phi\|_p \le 1$, both functions with mean value zero.

Modifying formula (4.1), we now construct linear operators U_n which provide good trigonometric approximation for functions f. We fix $r = 1, 2, \dots$ and let $R_n := R_{n,r}$ denote the best L_1 approximation to the Bernoulli spline $\mathcal{B}_r$ by trigonometric polynomials of degree $\le n-1$. The operator

$$U_n(f,x) := \frac{1}{\pi}\int_{\mathbb{T}} f^{(r)}(t)\, R_n(x-t)\,dt \tag{4.4}$$

assigns to each $f \in W_1^r(\mathbb{T})$ a trigonometric polynomial of degree $< n$. We shall see that the U_n has the desired approximation properties. We define $h_n := h_{n,r}$, $r = 1, 2, \dots$, $n = 1, 2, \dots$ by

$$h_n(x) := \begin{cases} \cos nx\,, & \text{if } r \text{ is even} \\ \sin nx\,, & \text{if } r \text{ is odd.} \end{cases} \tag{4.5}$$

Clearly the functions $h_{n,r}$ and $\mathcal{B}_r$, and the number r are of the same parity. It is convenient for the proof that follows to put $\mathcal{B}_1(0) := \frac{1}{2}(\mathcal{B}_1(0+)+\mathcal{B}_1(0-)) = 0$.

Theorem 4.1. (i) *For the polynomial R_n of degree $\le n-1$ that interpolates $\mathcal{B}_r$ at the $2n$ zeros of h_n, the difference $\mathcal{B}_r - R_n$ changes sign at each of these points, and has no other zeros.* (ii) *Moreover, R_n is the L_1-best approximation to $\mathcal{B}_r$ from $\mathcal{T}_{n-1}$, and*

$$\int_{\mathbb{T}} |\mathcal{B}_r(t) - R_n(t)|\,dt = \left|\int_{\mathbb{T}} \mathcal{B}_r(t)\,\operatorname{sign} h_n(t)\,dt\right|. \tag{4.6}$$

Proof. (i) The functions $\sin x, \dots, \sin(n-1)x$ are a Haar system on $[\delta, \pi-\delta]$, $\delta > 0$, and $1, \cos x, \dots, \cos(n-1)x$ have this property on $[0,\pi]$. The zeros of h_n are symmetric with respect to 0 on $[-\pi,\pi]$; they include 0 and π if r is odd. Therefore there is a unique trigonometric polynomial R_n whose parity matches that of $\mathcal{B}_r$ and h_n and which interpolates $\mathcal{B}_r$ at the zeros of h_n.

It follows that $\Delta_n := \Delta_{n,r} = \mathcal{B}_r - R_n$ has $2n$ zeros (including the zero at 0 when n is odd) and has the same parity as r. We have to show that (counting multiplicities) Δ_n cannot have more than $2n$ zeros. Assume that this is not the case. First let $r \ge 2$. Then Δ_n has at least $2n+2$ zeros on $\mathbb{T}$. By Rolle's theorem for $\mathbb{T}$, also $\Delta_n^{(r-2)} = \mathcal{B}_2 - S_n$ (where $S_n = R_n^{(r-2)} \in \mathcal{T}_{n-1}$ is even) has at least $2n+2$ zeros. It follows that $\Delta_n^{(r-1)} = \mathcal{B}_1 - T_n$, with odd T_n, has at least $2n+1$ continuous zeros on $\mathbb{T} \setminus \{0\}$.

For $r = 1$ we have the same conclusion. Besides the $2n-1$ known continuous zeros, $\Delta_{n,1} = \mathcal{B}_1 - T_n$ (where $T_n = R_n$) has an additional zero $x_0 \ne 0$. If $x_0 = \pi$, its multiplicity is ≥ 3; if $0 < |x_0| < \pi$, then $-x_0$ is also a zero. It follows in both cases that

$$(\mathcal{B}_1 - T_n)' = -\frac{1}{2} - T_n'$$

is an even trigonometric polynomial of degree $\le n-1$ with $2n$ zeros and the constant term $-\frac{1}{2} \ne 0$. This is absurd.

(ii) From Lemma 10.3 of Chapter 3 we see that the function sign h_n is orthogonal to all trigonometric polynomials of degree $\le n-1$. Since sign $(\mathcal{B}_r - R_n) = \varepsilon \operatorname{sign} h_n$, $\varepsilon = \pm 1$, also $\operatorname{sign}(\mathcal{B}_r - R_n)$ has this property. Then Corollary 10.2(ii) of Chapter 3 insures that R_n is the best approximation to $\mathcal{B}_r$ in L_1. Relation (4.6) also follows. □

The main result of this section is the following theorem of Favard [1937].

Theorem 4.2. *For $r = 1, 2, \dots$,*

$$E_{n-1}(B_\infty^r)_\infty = K_r\, n^{-r}\,, \qquad n = 1, 2, \dots\,. \tag{4.7}$$

The function $f_0(t) := K_r\, n^{-r}\mathcal{E}_r(nt/\pi)$ where $\mathcal{E}_r$ is the Euler spline, is extremal: it satisfies $E_{n-1}(f_0)_\infty = K_r\, n^{-r}$.

Proof. From the properties of Euler splines in §5 of Chapter 5, the function f_0 has mean value zero and satisfies $\|f_0^{(r)}\|_\infty = 1$ (so that $f_0 \in B_\infty^r$) and $f_0^{(r)} = \operatorname{sign} h_n$. Moreover,

$$\|f_0\|_\infty = f_0(0) = M := K_r\, n^{-r}\,.$$

By (5.3) (vi) of Chapter 5, at the points $t_k = k\pi/n$ $k = -n, \dots, n-1$, f_0 alternately takes the values $\pm M$. From Chebyshev's theorem on $\mathbb{T}$, its best uniform approximation from $\mathcal{T}_{n-1}$ is the zero function. Therefore

$$E_{n-1}(B_\infty^r) \ge E_{n-1}(f_0) = f_0(0) = K_r\, n^{-r}\,. \tag{4.8}$$

On the other hand, for each $f \in B_\infty^r$ of mean value zero, we have

$$\begin{aligned} |f(x) - U_n(f,x)| &= \frac{1}{\pi}\left|\int_{\mathbb{T}} f^{(r)}(t)\,(\mathcal{B}_r(x-t) - R_n(x-t))\,dt\right| \\ &\le \frac{1}{\pi}\int_{\mathbb{T}} |\mathcal{B}_r(t) - R_n(t)|\;dt = \frac{1}{\pi}\left|\int_{\mathbb{T}} \mathcal{B}_r(t)\operatorname{sign} h_n(t)\,dt\right| \\ &= \frac{1}{\pi}\left|\int_{\mathbb{T}} \mathcal{B}_r(t) f_0^{(r)}(t)\,dt\right| = f_0(0) = K_r\, n^{-r}\,. \end{aligned}$$

This yields that $E_{n-1}(f)_\infty \le K_r\, n^{-r}$. □

As a corollary, the error of L_1-approximation of $\mathcal{B}_r$ in L_1 is equal to

$$\frac{1}{\pi}\int_{\mathbb{T}} |\mathcal{B}_r - R_n|\;dt = M = K_r\, n^{-r}\,. \tag{4.9}$$

The theorem of Favard is one of the fundamental results of the approximation theory, and has important applications. As a first example, we estimate the errors of the trigonometric approximation of functions $f \in W^r(X)$, where X is any rearrangement invariant Banach function space on $\mathbb{T}$. Here $W^r(X)$ denotes the space of all functions f for which $f^{(r-1)}$ is absolutely continuous and $f^{(r)} \in X$. For example, $W^r(L_p) = W_p^r$.

Theorem 4.3. *If $f \in W^r(X)$, then*

$$E_{n-1}(f)_X \le \|f - U_n f\|_X \le K_r n^{-r} \|f^{(r)}\|_X . \tag{4.10}$$

Proof. The operator

$$V_n(g, x) := \frac{1}{\pi} \int_{\mathbb{T}} g(t) \Lambda_n(x - t)\, dt , \qquad \Lambda_n(t) = \mathcal{B}_r(t) - R_n(t) \tag{4.11}$$

is a convolution operator of the type of Corollary 4.6 of Chapter 2. Its norms both on L_1 and L_∞ are equal to $K_r n^{-r}$. By this corollary, V_n maps also X into itself with norm $\le K_r n^{-r}$. Since $f - U_n(f) = V_n(f^{(r)})$, we have

$$\|f - U_n(f)\|_X = \|V_n(f^{(r)})\|_X \le K_r n^{-r} \|f^{(r)}\|_X . \qquad \square$$

All of the constants K_m are bounded from above by $K_1 = \pi/2$. This allows us to improve significantly the approximation estimates of §2 making the constants independent of the order of the derivative. For example, with an *absolute constant* C,

$$E_{n-1}(f)_p \le C n^{-r} \omega \left(f^{(r)}, n^{-1}\right)_p , \qquad f \in W_p^r . \tag{4.12}$$

We can improve this, replacing L_p by X and $\omega = \omega_1$ by ω_k:

Theorem 4.4. *Let X be a rearrangement-invariant space in which translation is continuous. For each $k = 1, 2, \dots$ there is a constant C_k with the property*

$$E_{n-1}(f)_X \le C_k n^{-r} \omega_k \left(f^{(r)}, n^{-1}\right)_X , \qquad f \in W^r(X) , \tag{4.13}$$

where ω_k is the modulus of smoothness for the space X.

Proof. From the remark following Corollary 2.5, applied to $f^{(r)}$, there exists a trigonometric polynomial $S_{n-1} \in \mathcal{T}_{n-1}$ for which $\|f^{(r)} - S_{n-1}\|_X \le C_k \omega_k \left(f^{(r)}, \frac{1}{n}\right)_X$. Replacing C_k by $2C_k$ and using (2.15), we can assume that S_{n-1} has mean value zero; then $S_{n-1} = T_{n-1}^{(r)}$, for some $T_{n-1} \in \mathcal{T}_{n-1}$. The desired inequality then follows from (4.10):

$$\begin{aligned} E_{n-1}(f)_X = E_{n-1}(f - T_{n-1})_X &\le K_1 n^{-r} \|f^{(r)} - T_{n-1}^{(r)}\|_X \\ &\le \frac{\pi}{2} 2C_k n^{-r} \omega_k \left(f^{(r)}, n^{-1}\right)_X . \end{aligned} \qquad \square$$

According to Theorem 4.3, $E_{n-1}(B_p^r)_p \le K_r n^{-r}$. Favard has proved equality here for $p = \infty$, and Nikolskii [1946] established this also for $p = 1$:

Theorem 4.5 (Nikolskii). *One has*

$$E_{n-1}(B_1^r)_1 = K_r n^{-r} . \tag{4.14}$$

Proof. We have only to show that for each $\varepsilon > 0$, there is an $f \in W_1^r(\mathbb{T})$ satisfying $E_{n-1}(f) \ge M - \varepsilon$ where $M := K_r n^{-r}$. If f_0 and h_n are as in Theorem 4.2, then $h_n = \operatorname{sign} f_0^{(r)}$. Now $f_0(k\pi/n) = (-1)^k M$, $k \in Z$. We choose $\delta > 0$ so small that the intervals $I_k := [\pi k/n - \delta, \pi k/n + \delta]$ are disjoint and that $|f_0(x)| \ge M - \varepsilon$ on I_k, $k = 0, \dots, 2n-1$. Let $\phi(x) := (-1)^k/(4n\delta)^{-1}$ for $x \in I_k$, $:= 0$ elsewhere on $\mathbb{T}$. Then $\|\phi\|_1 = 1$ and ϕ has mean value zero. If f is the periodic r-th integral of ϕ, then $f \in W_1^r$. Since $f_0^{(r)}$ is orthogonal to all trigonometric polynomials of degree $< n$ and $\|f_0^{(r)}\|_\infty = 1$, from (1.8) of Chapter 3, we have

$$E_{n-1}(f)_1 \ge (-1)^r \int_{\mathbb{T}} f f_0^{(r)}\, dt = \int_{\mathbb{T}} f^{(r)} f_0\, dt = \int_{\mathbb{T}} \phi f_0\, dt \ge M - \varepsilon . \qquad \square$$

For a remarkable extension of Favard's theorems by Korneichuk see his book [A-1976] and §4 of Chapter 11.

§ 5. Improvement of Estimates

There is a standard method of improving the error estimates, such as Jackson's inequality (2.11), by replacing $\omega(f, t)$ by $\omega_r(f, t)$ or by a K-functional $K(f, t)$. This method disregards the value of constants and is due to Peetre, who often emphasized the importance of K-functionals in approximation. The theorems below (originated by Butzer and Scherer [1972], and given in the present form by DeVore and Popov [1988]) will be used often in this and the following chapters. The first of them will appear in a general setting (quasi-normed spaces, not necessarily linear sets Φ_n) for later use and in §9.

Let $\|\cdot\| := \|\cdot\|_X$ be the quasi-norm (or norm) on a complete linear space X. In view of Theorem 1.1 of Chapter 2, we can assume that

$$\|f + g\|^\mu \le \|f\|^\mu + \|g\|^\mu \tag{5.1}$$

for some $\mu := \mu(X)$. We shall approximate $f \in X$ by the elements of the subsets Φ_n, $n = 0, 1, \dots$, of X which are not necessarily linear. We assume that the sequence $\Phi := (\Phi_n)_0^\infty$ has the following properties:

$$(5.2)\quad \begin{cases} \text{(i)} & 0 \in \Phi_n \ ; \ \Phi_0 := \{0\}\,, \\ \text{(ii)} & \Phi_n \subset \Phi_{n+1}\,, \\ \text{(iii)} & a\Phi_n = \Phi_n \ \text{ for each } \ a \neq 0\,, \\ \text{(iv)} & \Phi_n + \Phi_n \subset \Phi_{cn}, \ \text{ with a constant } \ c := c(\Phi) \\ \text{(v)} & \cup \Phi_n \ \text{ is dense in } \ X\,, \\ \text{(vi)} & \text{Each } \ f \in X \ \text{ has a best approximation from } \ \Phi_n \\ & \text{for } \ n = 0, 1, \ldots . \end{cases}$$

If the sets are linear, one can take $c = 1$ in (iv).

Examples for the sets Φ_n are the trigonometric polynomials $\mathcal{T}_{n-1}$ on $\mathbb{T}$, the sets $\mathcal{P}_{n-1}$ of algebraic polynomials, or the sets $\mathcal{R}_{n-1}$ of rational functions of *degree* $\le n-1$ (that is, of functions $R_n = P_n/Q_n$, where $P_n, Q_n \in \mathcal{P}_{n-1}$) on an interval, with $c = 2$.

For all $f \in X$, we define the error of approximation of f by the Φ_n:

$$(5.3)\quad E_n(f) := E_n(f)_X := \inf_{\varphi \in \Phi_n} \|f - \varphi\|_X\,, \quad n = 1, 2, \ldots\,; \ E_0(f) = \|f\|_X\,.$$

We assume that there is a linear space $Y := Y_r$, $r > 0$, with a semiquasi-norm $|\cdot|_Y$ that is continuously embedded in X so that the following inequalities are valid for $n = 1, 2, \ldots$:

$$(5.4)\qquad E_n(f)_X \le C n^{-r} |f|_Y\,, \qquad f \in Y\,,$$

$$(5.5)\qquad |\varphi|_Y \le C n^r \|\varphi\|_X\,, \qquad \varphi \in \Phi_n\,.$$

Because of Theorem 1.1 of Chapter 2, we can assume that

$$(5.6)\qquad |f + g|_Y^\mu \le |f|_Y^\mu + |g|_Y^\mu\,.$$

where $\mu := \mu(Y)$ is the exponent of that theorem. Decreasing $\mu(X)$ or $\mu(Y)$, if necessary, we shall assume that they are equal.

The inequality (5.4) implies that the null space $\mathcal{N}$ of the seminorm $|\cdot|_Y$ is contained in Φ_n for all $n \ge 1$. We shall assume that $\Phi_1 = \mathcal{N}$ (this can always be accomplished by simply adding $\mathcal{N}$ to the sequence Φ). With Butzer and Scherer [1972], we call (5.4) a Jackson inequality and (5.5) a Bernstein inequality. Indeed, if $X = L_p(\mathbb{T})$, $Y_r = W_p^r(\mathbb{T})$, and $\Phi_n = \mathcal{T}_{n-1}$, $n = 1, 2, \ldots$ then (5.4) follows from (2.12) and (5.5) is Theorem 2.5 of Chapter 4.

In the following theorem (by Peetre) the K-functionals $K(f,t) := K(f,t;X,Y)$ of the pair X, Y (see §1 of Chapter 6) appear in a natural way.

Theorem 5.1. *Let X, Y, $r > 0$, and (Φ_n) be as stated above.*

(i) *If the Jackson inequality* (5.4) *is satisfied for $n = 1, 2, \ldots$, then*

$$(5.7)\qquad E_n(f)_X \le C K(f, n^{-r})\,, \qquad f \in X\,, \quad n = 1, 2, \ldots .$$

(ii) *If the Bernstein inequality* (5.5) *holds, then with $\mu = \mu(Y)$,*

$$(5.8)\quad K(f, n^{-r}) \le C n^{-r} \left\{ \sum_{k=1}^{n} [k^r E_k(f)]^\mu \frac{1}{k} \right\}^{1/\mu}, \qquad f \in X\,, \ n = 1, 2, \ldots .$$

Proof. (i) For each $g \in Y$, we have from (5.4),

$$E_n(f) \le C(\|f - g\|_X + E_n(g)) \le C(\|f - g\|_X + n^{-r}|g|_Y)$$

and (5.7) follows by taking an infimum over all $g \in Y$.

(ii) It will be sufficient to prove (5.8) for $n = 2^m$; for other values of n it follows from the monotonicity of $K(f,.)$. Let $\varphi_k \in \Phi_{2^k}$ be a best approximation to f : $\|f - \varphi_k\|_X = E_{2^k}(f)$, $k = 0, 1, \ldots$. Let $\psi_k := \varphi_k - \varphi_{k-1}$, $k = 1, 2, \ldots$, $\psi_0 := \varphi_0$. Then $\psi_k \in \Phi_{c2^k}$ by property (5.2) (iv). Using (5.1) we have

$$\|\psi_k\|_X^\mu \le \|f - \varphi_k\|_X^\mu + \|f - \varphi_{k-1}\|_X^\mu \le 2E_{2^{k-1}}(f)^\mu\,, \qquad k = 1, 2, \ldots .$$

Since $\varphi_m = \sum_{k=0}^m \psi_k$ and $|\psi_0|_Y = 0$, it follows from (5.5) that

$$(5.9)\quad \begin{aligned} K(f, 2^{-mr}) &\le \|f - \varphi_m\|_X + 2^{-mr}|\varphi_m|_Y \\ &\le E_{2^m}(f) + 2^{-mr} \left(\sum_{k=1}^m |\psi_k|_Y^\mu \right)^{1/\mu} \\ &\le E_{2^m}(f) + C2^{-mr} \left(\sum_{k=1}^m [2^{kr} \|\psi_k\|_X]^\mu \right)^{1/\mu} \\ &\le C2^{-mr} \left(\sum_{k=1}^m [2^{kr} E_{2^{k-1}}(f)]^\mu \right)^{1/\mu} . \end{aligned}$$

From the monotonicity of E_j, this last sum is less than a multiple of the right side of (5.8) (with $n = 2^m$). □

As an example, let $X = L_p(\mathbb{T})$, $Y = W_p^r(\mathbb{T})$, $1 \le p \le \infty$, $r = 1, 2, \ldots$, and $\Phi_n = \mathcal{T}_{n-1}$, $n = 1, 2, \ldots$ in which case one can take $\mu = 1$. Taking into account the equivalence of K-functionals and moduli of smoothness (Theorem 2.4 of Chapter 6), one sees that Theorem 2.3 of the present chapter is a special case of (5.7) while (3.1) follows from (5.8).

There exist similar theorems by Freud [1968] for bounded linear operators, instead of the operator of best approximation. We state them here for the special case when $X = L_p(A)$, $Y_r = W_p^r(A)$, for standard sets A.

Theorem 5.2. *Let $1 \le p \le \infty$, $r = 1, 2, \ldots$ and let $\delta_n > 0$, $\delta_n \to 0$. If $(U_n)_1^\infty$, $\|U_n\| \le M$ is a uniformly bounded sequence of linear operators on L_p and if for each $g \in W_p^r$*

$$\|g - U_n g\|_p \le \delta_n^r \left(\|g\|_p + \|g^{(r)}\|_p\right), \tag{5.10}$$

then for each $f \in L_p$, *we have*

$$\|f - U_n f\|_p \le C\left(\|f\|_p \delta_n^r + \omega_r(f, \delta_n)_p\right) \tag{5.11}$$

with C *depending only on* r *and* M. *If* $\|g\|_p$ *is absent in* (5.10), *then the term with* $\|f\|_p$ *can be omitted in* (5.11).

Proof. Let n be fixed. From the definition of the K-functional and (2.11) of Chapter 6, for each $\varepsilon > 0$, there is a $g \in W_p^r$ such that with all norms in L_p,

$$\|f - g\| + \delta_n^r \|g^{(r)}\| \le C\omega_r(f, \delta_n) + \varepsilon. \tag{5.12}$$

Hence, from (5.10)

$$\begin{aligned} \|f - U_n(f)\| &\le \|f - g\| + \|g - U_n(g)\| + \|U_n(g - f)\| \\ &\le (1 + M)\|f - g\| + (\|g\| + \|g^{(r)}\|)\delta_n^r \\ &\le C\left(\omega_r(f, \delta_n) + \varepsilon\right) + \|g\|\delta_n^r. \end{aligned} \tag{5.13}$$

From (5.12), $\|g\| \le \|f\| + C\omega_r(f, \delta_n) + \varepsilon \le C\|f\| + \varepsilon$. If we use this in (5.13) and let $\varepsilon \to 0$, we obtain (5.11). If $\|g\|$ does not appear in (5.10), then it will not appear on the right side of (5.13) and therefore $\|f\|$ can be omitted in (5.11). □

A variant of Theorem 5.2 gives pointwise estimates. Let $\delta_n(x) \ge 0$, $x \in A$.

Theorem 5.3. *Let* $r = 1, 2, \ldots$. *If* $(U_n)_1^\infty$, $\|U_n\| \le M$, *are bounded operators on* $C(A)$ *and for each* $g \in C^r(A)$,

$$|g(x) - U_n(g, x)| \le \|g^{(r)}\|_\infty \, \delta_n(x)^r, \qquad x \in A, \tag{5.14}$$

then for each $f \in C(A)$, *we have*

$$|f(x) - U_n(f, x)| \le C\omega_r\left(f, \delta_n(x)\right), \tag{5.15}$$

with C *depending only on* r *and* M.

Proof. Let $x_0 \in A$ and $\varepsilon > 0$ be fixed. We have again (5.12) with $\delta_n := \delta_n(x_0)$, and by the hypothesis (5.14),

$$|g(x_0) - U_n(g, x_0)| \le C\omega_r(f, \delta_n) + \varepsilon.$$

As in (5.13), this and (5.14) imply that for each $f \in C(A)$ and an arbitrarily small $\varepsilon > 0$,

$$|f(x_0) - U_n(f, x_0)| \le C\left(\omega_r(f, \delta_n) + \varepsilon\right). \qquad □$$

§ 6. Approximation by Algebraic Polynomials

An estimate of the error of approximation $E_n(f, A)$ of f by algebraic polynomials $P_n \in \mathcal{P}_n$ on A is contained in Theorem 4.2 of Chapter 6.

Theorem 6.1. *For* $f \in L_p(I)$, $I = [a, b]$, $1 \le p \le \infty$, *one has*

$$E_r(f, I)_p \le C_r \omega_r(f, |I|)_p, \qquad r = 1, 2, \ldots. \tag{6.1}$$

This is valuable for fixed r and for intervals I of small length $|I|$, for example, in spline approximation.

Quite different are estimates of $E_n(f, A)$ of Jackson type. They are interesting for large n, and for a fixed interval A, which we take to be $A := [-1, 1]$. Although the estimates of this type developed below are best possible of their kind, they have no true inverse theorems. The reason is that the endpoints ± 1 play, as Nikolskii [1946] was the first to recognize, a special role in this question. A function $f(x)$ in $C[-1, 1]$ can, in general, be better approximated if x is close to the endpoints. One can improve the estimates of this section to *local approximation theorems*, for which the upper bound for the deviation $|f(x) - P_n(x)|$ depends on the position of x in $[-1, 1]$. The rival theories of Dzyadyk-Timan and of Ditzian-Totik which allow this, will be the subject of Chapter 8.

We shall now give two "poor man's theorems" which disregard these finer points, and are justified by the relative simplicity of their proofs. The key inequality

$$E_n(f)_p \le C\omega(f, 1/n)_p \tag{6.2}$$

does not follow by the "standard substitution" $g(t) = f(\cos t)$ in the corresponding trigonometric inequality (2.7), except when $p = \infty$, since $g \in L_p(\mathbb{T})$ is equivalent to $\int_A |f(x)|^p (1 - x^2)^{-1/2}\, dx < \infty$, and not to $f \in L_p(A)$. To get (6.2) we use K-functionals and Theorem 2.4 of Chapter 6. This is equivalent to the following: we approximate $f \in L_p$ by a smooth function $g \in W_p^r$ which, at least on a part of interval A, is representable by a properly chosen integral operator (see (2.12) of Chapter 6). For $r = 1$, this integral reduces to the Steklov average.

Theorem 6.2. *For functions* $f \in W_p^r(A)$, $n > r$, $r = 0, 1, 2, \ldots$, $1 \le p \le \infty$, *the error of polynomial approximation satisfies*

$$E_n(f)_p \le Cn^{-r}\omega(f^{(r)}, 1/n)_p. \tag{6.3}$$

Proof. (a) First we establish that for each $f \in W_p^1(A)$, we have

$$E_n(f)_p \le \frac{\pi}{2(n+1)}\|f'\|_p. \tag{6.4}$$

We put $x = \cos t$ and $g(t) := f(\cos t)$; this function is even. It belongs to $W_p^1(\mathbb{T})$ because

$$\begin{aligned}\|g'\|_p^p &= \int_{\mathbb{T}} |g'(t)|^p\,dt = 2\int_0^\pi |g'(t)|^p\,dt \\ &= 2\int_{-1}^{+1} |\sin t\, f'(x)|^p \frac{dx}{\sin t} \le 2\|f'\|_p^p .\end{aligned}$$

For the trigonometric polynomial T_n of best approximation to g in the L_p norm, by (4.10) with $X = L_p$, $r = 1$, (or, with a less precise constant, from (2.7)),

$$\|g - T_n\|_p \le \frac{\pi}{2(n+1)}\|g'\|_p .$$

Since T_n is even, there is a polynomial $P_n \in \mathcal{P}_n$ with $P_n(\cos t) = T_n(t)$. We obtain $\|f - P_n\|_p \le 2^{-1/p}\|g - T_n\|_p$, and then (6.4).

(b) To derive (6.2), we employ (5.7) of Theorem 5.1 (or the trivial first lines of the proof of this theorem), regarding (6.4) as the required Jackson's condition (5.4). From Theorem 2.2 of Chapter 6, we get (6.2):

$$E_n(f)_p \le CK(f, 1/n; L_p, W_p^1) \le C\omega(f, 1/n)_p . \tag{6.5}$$

(c) Lastly, for $r = 1, 2, \ldots$, $f \in W_p^1$, $n \ge 1$,

$$E_n(f)_p \le Cn^{-1}E_{n-1}(f')_p . \tag{6.6}$$

Indeed, we have $E_n(f)_p \le Cn^{-1}\|f'\|_p$. If P_n' is the polynomial of degree $\le n-1$ of best approximation to f', and P_n is one of its integrals, then

$$E_n(f)_p = E_n(f - P_n)_p \le Cn^{-1}\|f' - P_n'\|_p = Cn^{-1}E_{n-1}(f')_p . \tag{6.7}$$

From (6.6) and (6.5) our theorem follows by iteration. □

To go further we once more use Theorem 5.1 of Peetre:

Theorem 6.3. *For each $r = 1, 2, \ldots$ there is a constant C_r such that for $f \in L_p[-1, 1]$,*

$$E_n(f)_p \le C_r\omega_r\left(f, n^{-1}\right)_p , \qquad n \ge r . \tag{6.8}$$

Proof. The inequality $E_n(f)_p \le Cn^{-r}\|f^{(r)}\|$ (which follows from (6.3)) is the Jackson inequality (5.4) of Theorem 5.1, if $\Phi_n := \mathcal{P}_n$, $X := L_p(A)$, $Y := W_p^r(A)$. We see that (6.8) follows from Theorem 5.1 and Theorem 2.4 of Chapter 6. □

There is not a complete inverse to (6.8), see Chapter 8. One can obtain however inverse statements for any proper subinterval $I := [-a, a]$, $0 < a < 1$, of A:

$$\omega_r\left(f, n^{-1}, I\right)_p \le Cn^{-r}\sum_{k=r}^n k^r E_k(f)_p \frac{1}{k} .$$

We leave the proof to the reader. For example, it follows that if $E_n(f)_\infty = \mathcal{O}(n^{-\alpha})$, then $f \in \mathrm{Lip}^*(\alpha, I)$ for each I. However on the whole interval A, this is no longer true.

Example. For the function

$$f(x) := \sum_{k=1}^\infty 2^{-\alpha k}C_{2^k}(x) , \qquad 0 < \alpha < 1 \tag{6.9}$$

(where C_m, $m = 1, 2, \ldots$, are Chebyshev polynomials), we prove that

$$E_n(f)_\infty \le Cn^{-\alpha} , \quad n = 1, 2, \ldots ; \quad f \notin \mathrm{Lip}(\beta, \infty) , \quad \beta > \frac{1}{2}\alpha , \tag{6.10}$$

so that (3.7) does not imply (3.8) for polynomial approximation.

Indeed, since $\|C_k\| = 1$ for all k, defining m by $2^{m-1} \le n < 2^m$, we get

$$E_n(f)_\infty \le E_{2^{m-1}}(f) \le \sum_{k\ge m} 2^{-\alpha k} \le C'2^{-\alpha m} \le Cn^{-\alpha} .$$

On the other hand, each term of the sum (6.9) attains its maximum at $x = 1$, while $y_N := \cos\frac{\pi}{2N}$ is a zero of C_N, $N := 2^n$. From the inequalities

$$\begin{cases} 1 - y_N = 2\sin^2\dfrac{\pi}{4N} \le \dfrac{1}{8}\pi^2 N^{-2} \\ f(1) - f(y_N) \ge 2^{-\alpha n}\left(C_N(1) - C_N(y_N)\right) = N^{-\alpha} . \end{cases}$$

It follows that $f \notin \mathrm{Lip}(\beta, \infty) \supset \mathrm{Lip}(\alpha, \infty)$.

We shall give now a useful expression of $E_{r-1}(f)_p$ on $A := [a, b]$ by means of the divided differences of f. The emphasis is upon the dependence of $E_{r-1}(f)_p$ on f and on $|A| = b - a$; the computation of the constant $c_{p,r}$ is a separate problem.

Theorem 6.4. *For each function $f \in C^1[a, b]$, $1 \le p \le \infty$, $r > 1$ there exist points $a < x_1 < \cdots < x_r < b$ and $a \le \xi \le b$ for which*

$$E_{r-1}(f, [a, b])_p = c_{p,r}(b - a)^{r+\frac{1}{p}}[x_1, \ldots, x_r, \xi]f \tag{6.11}$$

and

$$c_{p,r} := E_{r-1}\left(\frac{x^r}{r!}, [0, 1]\right)_p . \tag{6.12}$$

Proof. We recall that the polynomial $P \in \mathcal{P}_{r-1}$ of best L_p-approximation to a continuous function $g \in C[a, b]$ interpolates g at some r distinct points in (a, b) (see Theorem 10.8 of Chapter 3). For the function $g(t) = \frac{t^r}{r!}$ on $[0, 1]$, and its polynomial of best approximation, let these points be $0 < t_1^* < \cdots < t_r^* < 1$. Since $[t_1^*, \dots, t_r^*, t]g = g^{(r)}(\xi) = 1$, we have

$$c_{p,r} := E_{r-1}(g, [0,1])_p = \|(t - t_1^*) \cdots (t - t_r^*)\|_p = \min \|(t - t_1) \cdots ((t - t_r)\|_p$$

for any other points $t_i \in [0,1]$.

For the interval $[a,b]$, we define $x_i^* = a + (b-a)t_i^*$, then the product $Q^*(x) = (x - x_1^*) \cdots (x - x_p^*)$ satisfies

$$\|Q^*\|_p[a,b] = \min_{x_i} \|(x - x_i) \cdots (x - x_r)\|_p[a,b]$$

and

$$\|Q^*\|_p = \left(\int_a^b |Q^*|^p \, dx\right)^{1/p}$$
$$= (b-a)^{r+1/p} \left(\int_0^1 |(t - t_1) \cdots (t - t_r)|^p \, dt\right)^{1/p} = (b-a)^{r+1/p} c_{p,r} \, .$$

If $P \in \mathcal{P}_{r-1}$ interpolates the given function f at the points x_i^*, then, by the definition of the divided difference,

$$|f(x) - P(x)| \le Q^*(x) \, [x_1^*, \dots, x_r^*, x] \, f \, .$$

This divided difference is continuous (see §7 of Chapter 4). There exists a point $\xi^* \in [a,b]$ with the property

$$E_{r-1}(f)_p \le \|f - P\|_p = \|Q^*\|_p \, |[x_1^*, \dots, x_r^*, \xi^*] \, f| \, . \tag{6.13}$$

On the other hand, the polynomial of best approximation $\widetilde{P} \in \mathcal{P}_{r-1}$ to f in $L_p[a,b]$ interpolates f at some points $a < \widetilde{x}_1 < \cdots < \widetilde{x}_r < b$ (again by Theorem 10.8 of Chapter 3). With $\widetilde{Q}(x) = (x - \widetilde{x}_1) \cdots (x - \widetilde{x}_r)$, we get

$$\begin{aligned} E_{r-1}(f)_p &= \|f - \widetilde{P}\|_p = \|\widetilde{Q}(x)[\widetilde{x}_1, \dots, \widetilde{x}_r, x] f\|_p \\ &\ge \|\widetilde{Q}\|_p \min_x |[(\widetilde{x}_1, \dots, \widetilde{x}_r, x] f| \\ &\ge \|Q^*\|_p \left|[\widetilde{x}_1, \dots, \widetilde{x}_r, \widetilde{\xi}] f\right| \, . \end{aligned} \tag{6.14}$$

We transform $\widetilde{x}_i, \widetilde{\xi}$ into x_i^*, ξ^* continuously by means of formulas $x_i = \lambda \widetilde{x}_i + (1-\lambda)x_i^*$, $\xi = \lambda\widetilde{\xi} + (1-\lambda)\xi^*$. For properly chosen λ, $0 \le \lambda \le 1$, we shall have (6.11). □

Corollary 6.5 (Phillips [1979]). *For each $f \in C^r[a,b]$, $r \ge 1$, and some $c \in [a,b]$,*

$$E_{r-1}(f, [a,b])_p = c_{p,r}(b-a)^{r+1/p} \frac{f^{(r)}(c)}{r!} \, , \tag{6.15}$$

Proof. This follows from the property (e) of the divided differences of §7, Chapter 4. □

§ 7. Spline Approximation

For approximation by splines, there are estimates similar to those for trigonometric and algebraic polynomials. In this section, we derive fundamental estimates for the approximation error

$$s_{r,T}(f)_p := \operatorname{dist}\big(f, \mathcal{S}_r(T,I)\big)_p$$

of L_p approximation by splines from the Schoenberg space $\mathcal{S}_r(T,I)$. Here $I := [0,1]$ and $T := T_n := (t_j)$ consists of n basic knots, $0 < t_1 \le \cdots \le t_n < 1$, and auxiliary knots $t_{-r+1} \le \cdots \le t_0 := 0$ and $1 =: t_{n+1} \le \cdots \le t_{n+r}$. We shall be interested in the dependence of this error on the spacing of knots (or, possibly, on their number n). We give only some basic inequalities of this type. For a detailed exposition of the subject see Chapter 12.

Because of their small support, the B-splines N_j, $j = -r+1, \dots, n$, for $\mathcal{S}_r(T,I)$ are particularly useful for constructing good spline approximants. Consider for example the approximation of a function $f \in C(I)$ by the elements of $\mathcal{S}_r(T,I)$. By Theorem 3.1 of Chapter 5, each $S \in \mathcal{S}_r(T,I)$ has the representation $S = \sum_{j=-r+1}^n c_j N_j$. The value of S at a point x involves only the terms for which $N_j(x) \ne 0$. Thus, if S is to be a good approximation to f, the c_j should involve only the values of f in the support of N_j. For each $j = -r+1, \dots, n$ we select a point $\xi_j \in I \cap \operatorname{supp} N_j$. A natural choice is then $c_j = f(\xi_j)$. This leads to the linear operators:

$$L_T(f) := \sum_{j=-r+1}^{n} f(\xi_j) N_j \tag{7.1}$$

which map $C(I)$ onto $\mathcal{S}_r(T,I)$ for each given T.

The following simple lemma estimates the approximation error $\|f - L_T(f)\|_\infty$ in terms of "the mesh length" of T:

$$\delta := \delta_T := \max_{0 \le j \le n} (t_{j+1} - t_j) \, . \tag{7.2}$$

Lemma 7.1. *For $f \in C(I)$ and δ given by (7.2), we have*

$$s_{r,T}(f)_\infty \le \|f - L_T(f)\|_\infty \le r\omega(f,\delta)_\infty \, . \tag{7.3}$$

Proof. Since the N_j are non-negative, add to one, and since the support of N_j is the interval (t_j, t_{j+r}) of length $\le r\delta$, we have

$$|f(x) - L_T(f,x)| \le \sum_j |f(x) - f(\xi_j)| N_j(x) \le \omega(f, r\delta) \sum_j N_j(x) \le r\omega(f,\delta) \, .$$

□

We can go further and obtain estimates which take into account the differentiability of f. For this, we shall use a variant of the argument used in §2 and §6. If T is a knot sequence, we shall denote by T_0 the knot sequence which consists of the distinct points of T. To each of them we assign multiplicity one in T_0. It follows that $\delta := \delta_T = \delta_{T_0}$, $\mathcal{S}_r(T_0, I) \subset \mathcal{S}_r(T, I)$, and

$$s_{r,T}(f)_p \le s_{r,T_0}(f)_p \,. \tag{7.4}$$

The importance of (7.4) is that it allows us to derive estimates for $s_{r,T}$, by just considering the case when T consists only of simple knots. If the knots in T are all simple, each of the spaces $\mathcal{S}_k(T, I)$, $k \ge 1$ is defined.

Theorem 7.2. *Let $\delta := \delta_T$. For $r = 1, 2, \dots$, there is a constant C_r with the following properties:*
(i) *for each $k = 0, \dots, r-1$ and $f \in C^k(I)$, there exists an $S \in \mathcal{S}_r(T, I)$, with*

$$\|f - S\|_\infty \le C_r \delta^k \omega(f^{(k)}, \delta) \,, \tag{7.5}$$

(ii) *for each $f \in C(I)$, there exists an $S \in \mathcal{S}_r(T, I)$ so that*

$$\|f - S\|_\infty \le C_r \omega_r(f, \delta) \,. \tag{7.6}$$

Proof. In view of (7.4), we can assume that the knots in T are all simple. We prove (7.5) by induction on k. The case $k = 0$ follows from (7.3). We can therefore assume that $r > 1$ and that (7.5) holds for all r with k replaced by $k-1$. By our induction hypothesis, there is a spline $S_0 \in \mathcal{S}_{r-1}(T, I)$ which satisfies

$$\|f' - S_0\|_\infty \le C\delta^{k-1} \omega(f^{(k)}, \delta) \,. \tag{7.7}$$

Let S_1 be any integral of S_0. Then $S_1 \in \mathcal{S}_r(T, I)$, $S_0 = S_1'$, and (7.7) becomes

$$\|f' - S_1'\|_\infty \le C\delta^{k-1} \omega(f^{(k)}, \delta) \,. \tag{7.8}$$

We apply (7.3) to the function $f - S_1$ and obtain for some $S_2 \in \mathcal{S}_r$,

$$\|f - S_1 - S_2\|_\infty \le r\omega(f - S_1, \delta)_\infty \le r\delta \|f' - S_1'\|_\infty \le C_r \delta^k \omega(f^{(k)}, \delta) \,,$$

which proves (i).

For $k = r-1$, and $f \in C^r$, (7.5) gives the existence of a spline $S \in \mathcal{S}_r(T, I)$ with $\|f - S\|_\infty \le C_r \|f^{(r)}\|_\infty \delta^r$. Therefore, (7.6) follows from (7.5) and Theorem 5.1(i). □

Replacing $f(\xi_j)$ by $\frac{1}{|A_j|}\int_{A_j} f(x)\,dx$ with $A_j = [t_k, t_{k+1}]$ contained in $I \cap \operatorname{supp} N_j$, the above argument can be extended to cover the case $1 \le p < \infty$.

However, we shall take another approach which has the advantage of producing the approximant S to f by means of a linear operator, namely the quasi-interpolants $Q_T(f)$ of (4.12), Chapter 5. The operator Q_T is a bounded projection of each space $L_p(I)$, $1 \le p \le \infty$ onto $\mathcal{S}_r(T, I)$. We have the representation

$$Q_T(f)(x) = \sum_{j=-r+1}^{n} \gamma_j(f) N_j(x) \,, \tag{7.9}$$

where γ_j are certain linear functionals defined on $L_1(I)$. It is useful to note that Q_T provides a close to best approximation from $\mathcal{S}_r(T, I)$ to each $f \in L_p$. Indeed, since Q_T is uniformly bounded for all T and p, we get from Lebesgue's Lemma 4.1 of Chapter 2,

$$s_{r,T}(f)_p \le \|f - Q_T(f)\|_p \le C s_{r,T}(f)_p \,. \tag{7.10}$$

Theorem 7.3. *For a quasi-interpolant $Q_T(f)$ of order r and for each $f \in L_p(I)$, $1 \le p < \infty$, $f \in C(I)$, $p = \infty$, one has with δ defined by (7.2),*

$$\|f - Q_T f\|_p \le C_r \omega_r(f, \delta)_p \,. \tag{7.11}$$

Proof. We shall use the estimate (4.17) of Theorem 4.5 of Chapter 5. For the interval $I_j := [t_{j-r+1}, t_{j+r}] \cap I$ of that estimate and the midpoint ξ_j of I_j, we let P_j be the Taylor polynomial of f of degree $r-1$ at ξ_j. Then according to the Taylor remainder formula (5.7) of Chapter 2,

$$E_{r-1}(f, I_j)_p \le \frac{1}{(r-1)!} |I_j|^r \|f^{(r)}\|_p(I_j) \le C\delta^r \|f^{(r)}\|_p(I_j) \,. \tag{7.12}$$

Since each $x \in I$ appears in at most $2r-1$ of the I_j, we can use (7.12) in (4.17) of Chapter 5 to obtain

$$\|f - Q_T(f)\|_p(I) \le C_r \delta^r \|f^{(r)}\|_p(I) \,. \tag{7.13}$$

From (7.13), we derive (7.11) by means of Theorem 5.2. □

Direct estimates for spline approximation when r is small can already be found in the book of Ahlberg, Nilson, and Walsh [A-1967]. The estimate (7.5) has been proven by Freud and Popov [1969] using another technique. The use of quasi-interpolants for spline approximation was introduced by de Boor and Fix [1973].

We can also use quasi-interpolant operators to approximate derivatives. For our next theorem, we assume that T consists only of simple knots which are almost uniformly distributed: for some constant $C_0 > 0$, $C_0\delta \le |t_{j+1} - t_j| \le \delta$, $j = 0, \dots, n$.

Theorem 7.4. *Under the above assumptions, for $r = 1, 2, \dots$, and $f \in W_p^r(I)$, $1 \le p \le \infty$, one has for each $k = 0, \dots, r-1$,*

$$\|(f - Q_T(f))^{(k)}\|_p \le C s_{r-k,T}(f^{(k)})_p \,, \tag{7.14}$$

where C depends only on r and C_0.

Proof. We first prove by induction on k the representation

$$(7.15)\qquad \big(Q_T(f)\big)^{(k)} = \sum_{j=-r+k+1}^{n} \gamma_{j,k}(f^{(k)})N_{j,r-k}\,, \qquad k=0,\dots,r-1\,,$$

where the $\gamma_{j,k}$ are functionals on $L_1(I_{j,k})$, $I_{j,k} := [t_{j-k}, t_{j+r}] \cap I$, $j = -r+k+1,\dots,n$, given by functions $\varphi_{j,k}$ on $I_{j,k}$ by means of

$$(7.16)\qquad \gamma_{j,k}(g) = \int_{I_{j,k}} g(x)\varphi_{j,k}(x)\,dx\,, \qquad \|\varphi_{j,k}\|_\infty \le C|I_{j,k}|^{-1}\,.$$

For $k=0$, the bound (4.11) of Chapter 5 for the norms of the functionals $\gamma_{j,0} = \gamma_j$ on L_1 and the Riesz representation theorem show the existence of functions $\varphi_{j,0}$ which satisfy (7.15) and (7.16). Assume then that (7.15) and (7.16) have been established for some k. From formula (3.11) of Chapter 5, for the differentation of a B-spline series, we deduce that

$$(7.17)\qquad \begin{aligned} \big(Q_T(f)\big)^{(k+1)} &= \sum_{j=-r+k+2}^{n} a_j(f)N_{j,r-k-1}\,,\\ a_j(f) &= \frac{\gamma_{j,k}(f^{(k)}) - \gamma_{j-1,k}(f^{(k)})}{t_{j+r-k-1}-t_j}\,. \end{aligned}$$

The representation (7.16) of the $\gamma_{j,k}$ yields

$$(7.18)\qquad a_j(f) = \int_{I_{j,k+1}} f^{(k)}(x)\psi_j(x)\,dx$$

where $\psi_j(x) := (\varphi_{j,k}-\varphi_{j-1,k})/(t_{j+r-k-1}-t_j)$. Therefore, by our induction hypothesis, the ψ_j vanish outside of $I_{j,k+1}$ and satisfy $\|\psi_j\|_\infty \le C\delta^{-1}|I_{j,k+1}|^{-1}$. If we take $f(x) = x^k$ in (7.17), the left side is zero and therefore $a_j(f) = 0$. This shows that ψ_j has mean value zero on $I_{j,k+1}$. We define $\varphi_{j,k+1}(x) := \int_{t_j}^x \psi_j(u)\,du$. An integration by parts on the right side of (7.18) shows that (7.15) and (7.16) hold if k is replaced by $k+1$.

We next derive a bound for $\|(Q_T(f))^{(k)}\|_p$. We assume that $1 \le p < \infty$; a similar argument applies when $p=\infty$. From Hölder's inequality and (7.16), it follows that

$$(7.19)\qquad |\gamma_{j,k}(g)| \le C|I_{j,k}|^{-1/p}\|g\|_p(I_{j,k})\,.$$

Since $\sum N_{j,r-k} = 1$, we have, if p' is the conjugate exponent to p,

$$(7.20)\qquad \begin{aligned} \sum_{j=-r+k+1}^{n} |\gamma_{j,k}(g)|N_{j,r-k} &\le \Big(\sum_{j=-r+k+1}^{n} |\gamma_{j,k}(g)|^p N_{j,r-k}\Big)^{1/p}\Big(\sum_{j=-r+k+1}^{n} N_{j,r-k}\Big)^{1/p'}\\ &\le \Big(\sum_{j=-r+k+1}^{n} |\gamma_{r,k}(g)|^p N_{j,r-k}\Big)^{1/p}\,. \end{aligned}$$

With $g = f^{(k)}$, we take an $L_p(I)$ norm in (7.20) and use (7.19) to obtain

$$(7.21)\qquad \|\big(Q_T(f)\big)^{(k)}\|_p \le \Big(\sum_{j=-r+k+1}^{n} |\gamma_{j,k}(f^{(k)})|^p|I_{j,k}|\Big)^{1/p} \le C\|f^{(k)}\|_p\,,$$

where we used the fact that a point $x \in I$ appears in at most $r+k$ of the intervals $I_{j,k}$, $j = -r+k-1,\dots,n$.

Now to prove (7.14), let $S_0 \in \mathcal{S}_{r-k}(T,I)$ satisfy $\|f^{(k)} - S_0\|_p = s_{r-k,T}(f^{(k)})$. Then there is an $S \in \mathcal{S}_r(T,I)$ such that $S^{(k)} = S_0$. Since $Q_T(S) = S$, we have

$$(7.22)\qquad \|f^{(k)} - \big(Q_T(f)\big)^{(k)}\|_p \le \|f^{(k)} - S^{(k)}\|_p + \|\big(Q_T(f-S)\big)^{(k)}\|_p\,.$$

In view of (7.21) both terms on the right side of (7.22) do not exceed a multiple of $s_{r-k,T}(f^{(k)})_p$. □

In particular, if the knots of T_n are equally spaced, then $\delta =: \delta_n = 1/n$, and (7.11) yields

$$(7.23)\qquad s_{r,T_n}(f)_p \le C\omega_r\big(f, n^{-1}\big)_p\,, \qquad n = 1,2,\dots$$

for each function $f \in L_p$.

A different type of spline approximation results if it is possible to adjust the location of the knots of the spline to the singularities of the function. This is the *free knot spline approximation.* The *free knot spline space* $\Sigma_{n,r} := \Sigma_{n,r}(I)$, $I = [0,1]$, $r = 0,1,\dots$, $n = 1,2,\dots$, consists of all splines of order $\le r$ with arbitrary n knots $T : 0 < t_1 \le \dots \le t_n < 1$ (with possible repetitions). Thus, $\Sigma_{n,r}$ is the union of all spaces $\mathcal{S}_r(T,I)$ for arbitrary T. In other words, a spline S on I belongs to $\Sigma_{n,r}$ if and only if it has the B-spline representation

$$(7.24)\qquad S(x) = \sum_{j=1}^{n+r} a_j N(t_{j-r+1},\dots,t_j,x)$$

(with some auxiliary knots $t_{-r+1} \le \dots \le t_0 \le 0$, $1 \le t_{n+1} \le \dots \le t_{n+r}$).

We see that $\Sigma_{n,r}$ is a non-linear space of elements S that depend on $2n+r$ parameters: n knots t_j, $n+r$ coefficients a_j in (7.24). We define the error of $\Sigma_{n,r}$-approximation of a function $f \in L_p(I)$, $1 \le p < \infty$, $f \in C(I)$, $p=\infty$, by

$$(7.25)\qquad \sigma_{n,r}(f)_p := \operatorname{dist}(f, \Sigma_{n,r})_p\,.$$

An important variant of $\Sigma_{n,r}$ is the space $\Sigma^*_{n,r}$ of piecewise polynomials of order r with $n+1$ pieces. It is given by arbitrary breakpoints $T^* : 0 = t_0^* < t_1^* < \dots < t_{n+1}^* = 1$. On each interval $[t_j^*, t_{j+1}^*]$, the spline $S \in \Sigma^*_{n,r}$ is an arbitrary polynomial of degree $\le r-1$. Clearly, $\Sigma_{n,r} \subset \Sigma^*_{n,r} \subset \Sigma_{nr,r}$. It follows that the corresponding approximation errors are closely connected:

$$\sigma_{nr,r}(f)_p \le \sigma^*_{n,r}(f)_p \le \sigma_{n,r}(f)_p\,.$$

In contrast, the *fixed knot approximation error* $s_{r,T}(f)_p$, for any *fixed* sequence $(T_n)_1^\infty$ can be essentially larger than $\sigma_{n,r}(f)_p$.

Also for $\Sigma_{n,r}$, the splines with simple knots play a special role: each continuous spline $S \in \Sigma_{n,r}$ is uniformly approximable by these (for discontinuous splines this is true in the L_1 sense). Since all splines are linear combinations of truncated powers, it is sufficient to deal with the functions $\phi_k(x) := (x-a)_+^k$, $a \in I$, $k = 1, 2, \dots, r-1$. Then our claim follows at once from the formula

$$\begin{aligned}\phi_k(x) &= \frac{1}{r\dots(r-k+1)}\,\frac{d^{r-k}}{dx^{r-k}}\phi_n(x)\\ &= \frac{1}{r!}\lim_{h\to 0} h^{-r+k}\Delta_h^{r-k}(\phi_n, x)\end{aligned}$$

(see also §2 of Chapter 5). From this formula it follows also that the *errors of approximation from* $\Sigma_{n,r}$ *and from* $\widetilde{\Sigma}_{n,r}$ *are identical*. The space $\widetilde{\Sigma}_{n,r}$ consists of all splines of order $\le r$ with $\le n$ *simple* knots. The space $\widetilde{\Sigma}_{n,r}$ has the disadvantage that best approximations from it do not always exist.

We shall return to free knot spline approximation in Chapter 12. Here we give only an inverse theorem for this approximation by Burchard and Hale [1975] (a similar theorem for fixed knot approximation is a corollary). For its proof, we need also the space $\Sigma_{n,r}^{**}$, which consists of piecewise polynomials on I of degree $< r$ with $\le 2n$ breakpoints t_j which satisfy $|t_{j+1} - t_j| \le 1/n$ for all j. Then, $\Sigma_{n,r} \subset \Sigma_{n,r}^{**} \subset \Sigma_{2nr,r}$. As a consequence, for the error of approximation $\sigma_{n,r}^{**}$ by the elements of $\Sigma_{n,r}^{**}$, we have

$$\sigma_{2nr,r}(f)_p \le \sigma_{n,r}^{**}(f)_p \le \sigma_{n,r}(f)_p\,. \tag{7.26}$$

Theorem 7.6 (Burchard and Hale). *For* $1 \le p \le \infty$ *and a function* $f \in C^r(I)$, $I = [0,1]$, *the relation*

$$\sigma_{n,r}(f)_p = o(n^{-r}) \tag{7.27}$$

is possible only if f *is a polynomial of degree* $\le r-1$.

Proof. The relation (7.26) implies that $\sigma_{n,r}^{**}(f)_p = o(n^{-r})$. For a sequence $\varepsilon_n \to 0$, $\varepsilon_n > 0$, we select breakpoints $(t_j^{(n)})_1^{2n}$ such that $|t_{j+1}^{(n)} - t_j^{(n)}| \le 1/n$ and for $I_j^{(n)} := [t_j^{(n)}, t_{j+1}^{(n)}]$, we have

$$\sigma_{n,r}^{**}(f)_p^p \ge (1-\varepsilon_n)\sum_{j=0}^{2n} E_{r-1}(f, I_j^{(n)})_p^p\,, \qquad n = 1, 2, \dots\,. \tag{7.28}$$

We define $\gamma := r + \frac{1}{p}$, $\Lambda_n := \sum_{j=0}^{2n} E_{r-1}(f, I_j^{(n)})_p^{1/\gamma}$. Then by (6.15), for some $\xi_j^{(n)} \in I_j^{(n)}$, $j = 0, \dots, 2n$, and for $n \to \infty$,

$$\Lambda_n = c_{r,p}^{1/\gamma}\sum_{j=0}^{2n}|f^{(r)}(\xi_j^{(n)})|^{1/\gamma}|I_j^{(n)}| \to \text{Const}\int_0^1 |f^{(r)}|^{1/\gamma}\,dx\,.$$

We can estimate Λ_n by means of the sum in (7.28). By Hölder's inequality with the conjugate exponents γ/r and $p\gamma$,

$$\Lambda_n \le \left(\sum_0^{2n} 1\right)^{r/\gamma}\left(\sum_{j=0}^{2n} E_{r-1}(f, I_j^{(n)})_p^p\right)^{1/p\gamma} = (2n+1)^{r/\gamma}o(n^{-r/\gamma}) = o(1)\,.$$

It follows that $\int_0^1 |f^{(r)}|^{1/\gamma}\,dx = 0$. Hence, $f^{(r)}(x) = 0$ for all x and $f \in \mathcal{P}_{r-1}$. □

The assumption that $f \in C^r(I)$ is essential. For example, the powers $(x-a)_+^k$, $k = 0, \dots, r-1$, $a \in I$, can be identically approximated by the elements of $\Sigma_{n,r}$, once $n \ge r-k$.

§ 8. Approximation of Analytic Functions

A real or complex valued function f defined on $A := [-1, 1]$ is called *analytic* on A if there exists an analytic extension of f onto some open set G of the complex plane that contains A. We mean by this that there must exist on G a single-valued analytic function that coincides with f on A. If this extension exists, it is unique.

Examples of open sets that contain A are the open elliptic discs D_ρ, $\rho > 1$, bounded by the ellipses E_ρ with foci ± 1 and the sum of the half-axes ρ [see (2.2) of Chapter 4]. Moreover, for each open set $G \supset A$, we have $D_\rho \subset G$, if $\rho > 1$ is sufficiently close to 1.

From the properties of analytic functions, it follows that for each f analytic on A, there exists a $\rho_0 > 1$ characterized by the property that f has an analytic extension onto the disc D_{ρ_0}, but not onto any of the D_ρ for $\rho > \rho_0$. We must, however, admit the possibility $\rho_0 = +\infty$, which is realized for functions f analytic in the whole plane.

The error of approximation, $E_n(f)$ of an analytic function f by algebraic polynomials in the uniform norm on A, by (6.5), satisfies $E_n(f) = \mathcal{O}(n^{-\alpha})$ for each $\alpha > 0$. Actually, much more is true: $E_n(f)$ tends to zero not slower than a geometric progression q^n, $0 < q < 1$. The following theorem, due to Bernstein ([A-1952, vol. 1, pp. 41, 93]) even gives an asymptotic formula for $E_n(f)$. This is relation (8.1), given below, which is reminiscent of the relation

$$R^{-1} = \limsup \sqrt[n]{|c_n|}$$

between the radius R of the largest circle of analyticity of a function $g(z) = \sum_0^\infty c_n z^n$ and the coefficients c_n of its Taylor expansion.

Theorem 8.1. *A function* f*, defined on* A *is analytic on this interval if and only if* $\limsup \sqrt[n]{E_n(f)} < 1$*; and more exactly*

$$\limsup_{n\to\infty} \sqrt[n]{E_n(f)} = \frac{1}{\rho_0} \tag{8.1}$$

where ρ_0 is the number defined above. In particular, f has an analytic extension onto the whole plane if and only if

$$\limsup_{n\to\infty} \sqrt[n]{E_n(f)} = 0 .$$

Proof. (a) First we show that if f is analytic in D_ρ, $1 < \rho \le +\infty$, then

$$\limsup_{n\to\infty} \sqrt[n]{E_n(f)} \le \rho^{-1} . \tag{8.2}$$

We begin by expanding f on A into a series of Chebyshev polynomials:

$$f(x) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k C_k(x) . \tag{8.3}$$

To obtain this, we note that $f(\cos t)$ is an even, 2π-periodic function with a continuous derivative, and therefore has a convergent Fourier series:

$$f(\cos t) = \frac{1}{2}a_0 + \sum_{k=1}^{\infty} a_k \cos kt , \quad a_k = \frac{1}{\pi}\int_{-\pi}^{\pi} f(\cos t)\cos kt\, dt .$$

Substituting $\cos t = x$, we obtain (8.3). Since $|C_n(x)| \le 1$ on $[-1,+1]$, we have from (8.3)

$$E_n(f) \le \sum_{k=n+1}^{\infty} |a_k| . \tag{8.4}$$

We see that we must estimate $|a_k|$. The substitution $z = e^{it}$ in the integral for a_k gives the line integral along the circle $\Gamma : |z| = 1$:

$$a_k = \frac{1}{\pi i}\int_\Gamma f\left(\frac{z+z^{-1}}{2}\right)\frac{z^k + z^{-k}}{2}\,\frac{dz}{z} . \tag{8.5}$$

We take a ρ_1 with $1 < \rho_1 < \rho$ and consider the function

$$g(z) = f\left(\frac{z+z^{-1}}{2}\right)$$

in the closed annulus R bounded by the circles $\Gamma_1 : |z| = \rho_1^{-1}$ and $\Gamma_2 : |z| = \rho_1$. We know (§2 of Chapter 4) that if z is on either of the circles $|z| = \sigma$ or $|z| = \sigma^{-1}$, then $w = \frac{1}{2}(z+z^{-1})$ is on the ellipse E_σ. Since $f(w)$ is analytic in D_ρ, it follows that $g(z)$ is analytic in R. In order to obtain a good estimate of a_k, we now change the path of integration. We split the integral (8.5) into two parts; the integral containing z^{-k} is taken over the circle Γ_2, and for the integral with z^k, over the circle Γ_1. Thus

$$a_k = \frac{1}{2\pi i}\int_{\Gamma_1} g(z) z^{k-1}\,dz + \frac{1}{2\pi i}\int_{\Gamma_2} g(z) z^{-(k+1)}\,dz . \tag{8.6}$$

Let M be the maximum of the absolute value of $f(w)$ on the $\overline{D}_{\rho_1}$. Then the absolute value of the first integral does not exceed

$$\frac{1}{2\pi}M\left(\frac{1}{\rho_1}\right)^{k-1} 2\pi\rho_1^{-1} = M\rho_1^{-k} .$$

In the same way, the second integral is majorized by $M\rho_1^{-k}$, and we get $|a_k| \le 2M\rho_1^{-k}$. Then, by (8.4), we have

$$E_n(f) \le 2M \sum_{k=n+1}^{\infty} \rho_1^{-k} = \frac{2M}{\rho_1 - 1}\rho_1^{-n} = M_1\rho_1^{-n} . \tag{8.7}$$

Hence, $\sqrt[n]{E_n(f)} \le M_1^{1/n}\rho_1^{-1}$, and we obtain $\limsup \sqrt[n]{E_n(f)} \le \rho_1^{-1}$. This gives (8.2), since ρ_1 can be taken arbitrarily close to ρ.

(b) Conversely, let (8.2) hold for some ρ, $1 < \rho \le +\infty$. We show that f has an analytic extension onto each elliptic disc D_{ρ_1}, $1 < \rho_1 < \rho$ (and therefore onto D_ρ). Let $\rho_1 < \rho_2 < \rho$; then $\sqrt[n]{E_n(f)} \le \rho_2^{-1}$ for $n \ge n_0$ with a sufficiently large n_0. Writing

$$f(x) = P_0(x) + \sum_{n=0}^{\infty}[P_{n+1}(x) - P_n(x)] , \tag{8.8}$$

where the P_n are the polynomials of best approximation for f, we have

$$|P_{n+1}(x) - P_n(x)| \le 2E_n(f) \le 2\rho_2^{-n} , \qquad -1 \le x \le 1 ,\ n \ge n_0 .$$

Using the inequality (2.8) of Chapter 4, we can estimate this difference for all z in D_{ρ_1}:

$$|P_{n+1}(z) - P_n(z)| \le 2\rho_1\left(\frac{\rho_1}{\rho_2}\right)^n , \qquad z \in D_{\rho_1} ,\ n \ge n_0 .$$

Hence, the series (8.8), with x replaced by z, converges uniformly on D_{ρ_1}. Its sum is analytic on D_{ρ_1} and provides the desired analytic extension of f.

It follows from (a) and (b) that the largest elliptic disc D_ρ in which f is analytic is D_{ρ_0}, with ρ_0 given by (8.1). □

We can apply the expansion (8.3) to estimate the error of approximation of many entire functions. We use the following remark. *If the coefficients a_k in* (8.3) *satisfy*

$$\sum_{k=n+1}^{\infty} |a_k| \le \varepsilon_n |a_n| , \qquad \varepsilon_n \to 0 \ \textit{for}\ n \to \infty , \tag{8.9}$$

then for the uniform approximation on $A = [-1,1]$,

$$E_n(f) = |a_{n+1}|\bigl(1 + o(1)\bigr) . \tag{8.10}$$

Indeed, we take $P_n = \frac{a_0}{2} + \sum_{k=1}^n a_k C_k$ and have

$$f(x) - P_n(x) = a_{n+1}\big[C_{n+1}(x) + \delta_n(x)\big],$$

where $\delta_n(x) \to 0$ uniformly as $n \to \infty$. On A, the polynomials C_{n+1} have $n+2$ extrema ± 1 with alternating signs, consequently the difference $f(x) - P_n(x)$ will have at least $n+2$ extrema with changing signs and with absolute values $|a_{n+1}|(1+o(1))$. Thus, (8.10) follows from Theorem 5.2 of Chapter 3.

For an example, we consider $f(z) = e^{az}$, $a > 0$. We take both integrals (8.6) over $|z| = 1$. The substitution $z \to z^{-1}$ shows that they are equal. The second integral gives the coefficient of z^k in the Laurent expansion of g. To find this coefficient, we write

$$g(z) = \exp\left(\frac{a}{2}(z+z^{-1})\right) = \sum_{p=0}^{\infty} \frac{1}{p!}\left(\frac{a}{2}\right)^p (z+z^{-1})^p,$$

expand into powers of z and retain only terms with z^k. These will appear if $p = k + 2\nu$, $\nu = 0, 1, \ldots$. Hence, for $k = 0, 1, \ldots$

$$\begin{aligned} a_k &= 2\sum_{\nu=0}^{\infty} \frac{1}{(k+2\nu)!}\left(\frac{a}{2}\right)^{k+2\nu}\binom{k+2\nu}{\nu} \\ &= 2\sum_{\nu=0}^{\infty} \frac{(a/2)^{k+2\nu}}{\nu!(k+\nu)!}. \end{aligned}$$

It follows that

$$a_k = \frac{a^k}{2^{k-1}k!}\big(1+o(1)\big). \tag{8.11}$$

We see that the estimate (8.9) holds for the coefficients of e^{az}, hence from (8.10), we obtain:

Theorem 8.2 (Bernstein [A-1952, vol. 1, pp. 41, 93]). *In the uniform norm on* $[-1,1]$

$$E_n(e^{ax}) = \frac{2}{(n+1)!}\left(\frac{a}{2}\right)^{n+1}\big(1+o(1)\big). \tag{8.12}$$

Some important classes of entire functions can be characterized by means of their errors of approximation. Let

$$f(z) = \sum_{k=0}^{\infty} a_k z^k. \tag{8.13}$$

We define

$$M(r,f) := \max_{|z|\le r} |f(z)|. \tag{8.14}$$

An entire function f is said to be *of order not exceeding* $p \ge 0$, (we write then $f \in \Omega(p)$) if for each $\varepsilon > 0$ and a suitable constant $C := C(\varepsilon)$,

$$M(r,f) \le Ce^{r^{p+\varepsilon}}, \qquad r > 0. \tag{8.15}$$

(The order p_0 of f is the smallest p for which (8.15) is satisfied.)

Theorem 8.3 (Varga [1968]). *An entire function f belongs to the class $\Omega(p)$ if and only if on $[-1,1]$,*

$$E_n(f) \le Cn^{-n/(p+\varepsilon)} \quad \textit{for all} \quad \varepsilon > 0. \tag{8.16}$$

Proof. (a) The coefficients a_n of the Taylor expansion of f can be estimated by means of Cauchy's inequality and (8.15). For each $\lambda > 0$ we obtain

$$|a_n| \le r^{-n} M(r,f) \le Cr^{-n} e^{r^{p+\lambda}}. \tag{8.17}$$

We select $r = n^{1/(p+\lambda)}$; this gives $|a_n| \le Ce^n n^{-n/(p+\lambda)}$. If $s_n(z)$ is the n-th partial sum of (8.13), then with $\lambda = \frac{1}{2}\varepsilon$,

$$E_n(f) \le \max_{|x|\le 1} |f(x) - s_n(x)| \le C \sum_{k=n+1}^{\infty} e^k k^{-k/(p+\lambda)} \le C_1 n^{-n/(p+\varepsilon)}.$$

(b) Conversely, if (8.16) is satisfied, we can use (8.8) and have

$$|P_{n+1}(x) - P_n(x)| \le 2E_n(f) \le Cn^{-n/(p+\varepsilon)}, \qquad -1 \le x \le 1.$$

The smallest semi-axis of the ellipse D_ρ, $\rho > 1$ has length $\rho - \frac{1}{\rho}$, hence, the disc $|z| \le r$ is contained in D_{r+1}. From Theorem 2.2 of Chapter 4,

$$|P_{n+1}(z) - P_n(z)| \le 2Cn^{-n/q}(r+1)^n, \qquad |z| \le r,$$

where $q := p + \varepsilon$. Therefore on the disc $|z| \le r$,

$$|f(z)| \le C\sum_{n=1}^{\infty} \phi(n), \tag{8.18}$$

where $\phi(u) := u^{-u/q}(r+1)^u$. Taking the logarithmic derivative of ϕ, we see that the maximum of $\phi(u)$ is achieved when $u = u_0 := (r+1)^q/e$, and this maximum is, for large r, $\le C\exp r^{q+\varepsilon}$. On the other hand, for $u \ge u_1 := r^{q+\varepsilon}$ and large r, $\phi(u+1)/\phi(u) \le \frac{1}{2}$. It follows from (8.18) that for all large r

$$M(r,f) \le C(u_1+1)\exp r^{q+\varepsilon} \le e^{r^{p+3\varepsilon}}.$$

Since $\varepsilon > 0$ was arbitrary, $f \in \Omega(p)$. □

§ 9. Approximation Spaces

One can define spaces of functions by their approximation properties. Such spaces are provided by a collection of functions with common upper bounds for the errors of approximation. They have been studied in various contexts, by Besov in his Theorem 9.2 of this section about trigonometric approximation, by Butzer and Scherer [1972] and Peetre and Spaar [1972] for their relation to the interpolation of operators, and by Lorentz [1966] for the computation of metric entropy of sets of functions. Moreover, the results of the present section will be useful for the characterization of the approximation spaces for free knot splines in Chapter 12.

In the setting of §5, we approximate the elements f of a quasi-normed linear space X by elements of its subsets Φ_n, $n = 0, 1, \ldots$. We assume that the Φ_n have the properties (5.2), (but are not necessarily linear.) If $\delta := (\delta_n)_1^\infty$ is a sequence of positive numbers decreasing to zero, and $\mathbf{E}(f) := (E_{n-1}(f))_1^\infty$ is the sequence of errors of approximation

$$E_n(f) := \inf_{\varphi\in\Phi_n} \|f - \varphi\|, \qquad n = 1, 2, \ldots, \ E_0(f) := \|f\|,$$

we can define the approximation space $A(\delta, X)$ to consist of all of $f \in X$ that satisfy $E_{n-1}(f) = \mathcal{O}(\delta_n)$.

To obtain important special approximation spaces, we take $\delta_n := n^{-\alpha}$, $n = 1, 2, \ldots$, and apply to the sequence $E_{n-1}(f)$ a discrete α, q norm of (3.10) of Chapter 2. In this way we get the space $A_q^\alpha := A_q^\alpha(X) := A_q^\alpha(X, \Phi)$ which consist of all $f \in X$ for which the following quantity is finite for $\mathbf{E}(f)$:

$$\|f\|_{A_q^\alpha} := \|\mathbf{E}(f)\|_{\alpha,q} = \begin{cases} \left(\sum_{n=1}^{\infty} [n^\alpha E_{n-1}(f)]^q \dfrac{1}{n}\right)^{1/q}, & 0 < q < \infty \\ \sup_{n\ge1} [n^\alpha E_{n-1}(f)], & q = \infty. \end{cases} \tag{9.1}$$

If the term $n = 1$ is omitted in (9.1), we obtain the seminorm $|f|_{A_q^\alpha}$ of A_q^α. If $\|f\|_{A_q^\alpha} = 0$, then $E_0(f) = \|f\| = 0$, and $f = 0$. From (iii) of (5.2) we derive that $\|\cdot\|$ of (9.1) is homogeneous; from (iv), that $E_{cn}(f+g) \le C(E_n(f) + E_n(g))$. In this way, we see that (9.1) is a quasinorm for A_q^α, and a norm if X is a normed space, if the Φ_n are linear, and $1 \le q \le \infty$. It is easy to see that the spaces A_q^α decrease with increasing α and for fixed α with decreasing q. Therefore we have the continuous embeddings

$$A_q^\alpha \subset A_{q_1}^\beta, \quad \text{if} \quad \alpha > \beta \quad \text{or if} \quad \alpha = \beta \quad \text{and} \quad q \le q_1. \tag{9.2}$$

It is useful to mention that since the $E_n(f)$ are monotonically decreasing, we have the weak equivalences:

$$|f|_{A_q^\alpha} \sim \left\{ \begin{matrix} \left(\sum_{n=0}^{\infty} [2^{n\alpha} E_{2^n}(f)]^q\right)^{1/q}, & 0 < q < \infty \\ \sup_{n\ge0} [2^{n\alpha} E_{2^n}(f)], & q = \infty. \end{matrix} \right\} \tag{9.3}$$

Let $Y := Y_r$ be a linear space, with a quasinorm $|\cdot|_Y$ or a quasi-seminorm, which is continuously embedded in X. In view of Theorem 1.1 of Chapter 2, we can assume that (5.6) holds with $\mu = \mu(Y)$ being the exponent of that theorem. If the Jackson (5.4) and Bernstein (5.5) inequalities are valid for the pair X, Y, then we can characterize completely the approximation spaces A_q^α by means of the interpolation spaces $(X, Y)_{\theta,q}$ of §7, Chapter 6.

Theorem 9.1. *If both the Jackson and the Bernstein inequalities* (5.4) *and* (5.5) *hold for the spaces X, Y, then for $0 < \alpha < r$ and $0 < q \le \infty$ we have*

$$A_q^\alpha = (X, Y)_{\alpha/r,q}. \tag{9.4}$$

Proof. Let $Z := (X, Y)_{\alpha/r,q}$. We know that $\|f\|_Z = \|f\|_X + \rho(f)_{\theta,q}$ where $\rho(f)_{\theta,q}$ is equivalent to the discrete quasinorms (7.4), (7.5) of Chapter 6 with $\theta = \alpha/r$. The first of these and (5.7) yield the inequality $|f|_{A_q^\alpha} \le C\|f\|_Z$. In the opposite direction, by means of (7.5) of Chapter 6 we obtain

$$\rho(f)_{\theta,q} \sim \left(\sum_{n=0}^{\infty} [2^{n\alpha} K(f, 2^{-nr})]^q\right)^{1/q} \sim \|(b_k)\|_{\alpha,q}$$

where $b_k := K(f, 2^{-kr})$, $k = 0, 1, \ldots$, $b_k := 0$, for $k \le 0$. Now for $a_k := E_{2^k}(f)$, $k = 0, 1, \ldots$, $a_k := 0$, $k < 0$, we have from (5.8) and the monotonicity of $E_n(f)$,

$$b_k \le C 2^{-kr} \left(\sum_{j=-\infty}^{k} [2^{jr} a_j]^\mu\right)^{1/\mu}.$$

The conditions of the remark following Lemma 3.4 of Chapter 2 are satisfied and we obtain from this the inequality $\|(b_k)\|_{\alpha,q} \le C\|(a_k)\|_{\alpha,q}$ for all $\alpha < r$. It follows that, $\|f\|_Z \le C\|f\|_{A_q^\alpha}$. □

We can apply Theorem 9.1 to the case of trigonometric polynomial approximation. As noted earlier, the Jackson and Bernstein inequalities are valid for $X = L_p(\mathbb{T})$, $Y = W_p^r(\mathbb{T})$, $1 \le p \le \infty$, $r = 1, 2, \ldots$ and $\Phi_n = \mathcal{T}_{n-1}$. Since by Example 3 of §7 of Chapter 6, the interpolation spaces for L_p, W_p^r are the Besov spaces $(L_p, W_p^r)_{\alpha/r,q} = B_q^\alpha(L_p)$, we obtain as an application of Theorem 9.1 the following result of Besov:

Theorem 9.2. *For approximation by trigonometric polynomials $\Phi := (\mathcal{T}_n)$ in $L_p(\mathbb{T})$, $1 \le p < \infty$, or $C(\mathbb{T})$, $p = \infty$, we have for each $0 < \alpha < \infty$ and $0 < q \le \infty$,*

$$A_q^\alpha(L_p) = B_q^\alpha(L_p). \tag{9.5}$$

According to Theorem 9.1, in order to identify, for given X and Φ, the approximation spaces A_q^α, $0 < \alpha < r$, it is enough to find a space Y for which the Jackson and Bernstein inequalities are valid. It turns out that there are always such spaces Y, in fact, the space A_q^r is of this type for all $0 < q \le \infty$.

Theorem 9.3 (DeVore and Popov [1986]). *Let X be a space with a sequence Φ of sets that have the properties* (5.2). *Then for $0 < q \le \infty$, $0 < r < \infty$, the space $Y := Y_r := A_q^r$ satisfies the Jackson and Bernstein inequalities* (5.4) *and* (5.5). *Moreover, for $0 < \alpha < r$ and $0 < q_1 \le \infty$,*

$$(X, Y)_{\alpha/r, q_1} = A_{q_1}^{\alpha} . \tag{9.6}$$

Proof. The Jackson inequality (5.4) follows from the embedding $Y \subset A_\infty^r$. On the other hand, if $\varphi \in \Phi_n$, then $E_m(\varphi) = 0$, $m \ge n$, and $E_m(\varphi) \le \|\varphi\|_X$ for $0 \le m < n$. Hence, for $0 < q < \infty$,

$$|\varphi|_Y^q \le \sum_{m=2}^{n} [m^r \|\varphi\|_X]^q \frac{1}{m} \le C n^{rq} \|\varphi\|_X^q ,$$

with a similar inequality for $q = \infty$. In view of Theorem 9.1, we have established (9.6). □

An important consequence of (9.6) is that the spaces A_q^α, $\alpha > 0$, $0 < q \le \infty$, are invariant under the formation of θ, q spaces: for $\alpha_0, \alpha_1 > 0$, $0 < q_0, q_1, q \le \infty$ and $0 < \theta < 1$,

$$\left(A_{q_0}^{\alpha_0}, A_{q_1}^{\alpha_1}\right)_{\theta,q} = A_q^\alpha , \quad \text{where} \quad \alpha := (1-\theta)\alpha_0 + \theta\alpha_1 . \tag{9.7}$$

Indeed, with $r > \alpha_0, \alpha_1$, relation (9.7) is an immediate corollary of (9.6) and the reiteration Theorem 7.3 of Chapter 6.

§ 10. Problems

10.1. For the Fejér means $\sigma_n(f)$ of f, the following are equivalent:

$$\|f - \sigma_n(f)\|_\infty \le \text{Const}\, n^{-\alpha}, \quad \text{and} \quad f \in \operatorname{Lip}\alpha,\ 0 < \alpha < 1 .$$

(Bernstein; thus, he almost proved Jackson's theorem.)

10.2. If $f \in C^1(\mathbb{T})$, then $nE_n(f) \to 0$; if $\sum E_n(f) < +\infty$, then $f \in C^1(\mathbb{T})$.

10.3. If $f \in C^1(\mathbb{T})$ has mean value zero and $|f'(t)| \le 1$, then $|f(t)| \le \pi/2$ (H. Bohr).

10.4. If for the error $E_n(f)$ of uniform trigonometric approximation to f,

$$\sum_{n=1}^{\infty} n^{-1/2} E_n(f) < +\infty ,$$

then the Fourier series of f converges absolutely (Bernstein).

10.5. Show that $Z \ne \operatorname{Lip} 1$ by examining the function $\psi(t) = \sum_1^\infty n^{-2} \sin nt$ on $\mathbb{T}$ (one has $E_n(\psi) \le Cn^{-1}$, but $\psi'(t) \to \infty$ as $t \to 0$).

10.6. A function f on $[-1,1]$ belongs to the class Lip 1 if and only if for some sequence of polynomials $P_n(x)$, $|f(x) - P_n(x)| \le Cn^{-1}$ and $|P_n'(x)| \le C_1$ (Dzyadyk).

10.7. Compute the constant $c_{p,r}$ of (6.12) for $p = \infty$, $p = 1$ and $p = 2$, and all r.

10.8. If $r = 1, 2, \dots$, T is a knot sequence on $I = [0,1]$, and $\delta(x) := \delta_T(x) := |I \cap [t_{i-r+1}, t_{i+r}]|$, $x \in [t_i, t_{i+1}]$, $t_{i+1} \ne t_i$, then $|f(x) - Q_T(x)| \le C_r \omega_r(f, \delta(x))$, $x \in I$.

§ 11. Notes

11.1. As a proof of lack of communication in mathematics, it is interesting to note that Akhiezer and Krein [1937], who used the first part of Favard's paper [1937] but did not know of the second, also proved the basic theorem of Favard, Theorem 4.2.

11.2. We have often proved direct theorems of approximation by constructing a uniformly bounded sequence (U_n) of linear operators whose range is the space of polynomials of degree $\le n$. The question arises whether it is possible to construct once and for all a universal sequence (U_n) which provides the error estimates for all orders of smoothness. Dahmen and Görlich [1974] have shown that this is not the case. For example, there can be no such sequence which provides the Jackson estimates for all r and also gives estimates for analytic functions as in §8.

11.3. Simultaneous approximation of functions $f \in C^r(\mathbb{T})$ and of its derivatives by trigonometric polynomials $T_n \in \mathcal{T}_n$, in the case when both n and r are large, has been studied by Czipszer and Freud [1958], whose results have been improved by Garkavi [1960]. For $f \in C^r(\mathbb{T})$ that is not a trigonometric polynomial, we define

$$\mathcal{E}_{n,r}(f) := \inf_{T_n \in \mathcal{T}_n} \max_{k=0,\dots,r} \left(\|f^{(k)} - T_n^{(k)}\| / E_n(f^{(k)})\right) .$$

If $m := \min(n, r)$ is sufficiently large, then for all $f \in C^r$, $\mathcal{E}_{n,r}(f) \le \frac{4}{\pi^2} \log m + C_1$, and for some $f \in C^r$, $\mathcal{E}_{n,r} \ge \frac{4}{\pi^2} \log m - C_2 \log\log\log m$.

11.4. The following theorem (Bernstein [A-1952, vol. 1, p. 63], see also Lorentz [A-1986, p. 38]) is often useful:

Theorem. *If f, g are two functions with continuous derivatives of order $n+1$ on $[a,b]$ and if*

$$\left|f^{(n+1)}(x)\right| \le g^{(n+1)}(x), \qquad a \le x \le b ,$$

then

$$E_n(f)_\infty \le E_n(g)_\infty .$$

Thus, if $|f^{(n+1)}(x)| \le M$ on $[-1,1]$, then $E_n(f)_\infty \le M 2^{-n}/(n+1)!$; in particular, $E_n(e^x)_\infty \le e 2^{-n}/(n+1)!$ (Compare Theorem 8.2.)

Chapter 8. Influence of Endpoints in Polynomial Approximation

§ 1. Introduction

We have seen in §6 of Chapter 7 that in order to obtain best possible theorems for the polynomial approximation of a function $f(x)$ on an interval – which we can assume to be $A := [-1,1]$ – one has to take into account the position of the point x in A. One can do this in several ways. One can use, of course, weighted approximation, in the spaces $L_{w,p}$ with weight w.

There are two theories, that are specifically approximation-theoretic. Firstly, one can assume certain natural smoothness properties of the function f, such as $f \in \operatorname{Lip}\alpha$, and then look how this assumption influences its approximation properties. Secondly, one can prescribe the approximation error, for example require it to be $\le Cn^{-\alpha}$, and then try to find smoothness properties of f which guarantee this. In the first theory (that could be called the Dzyadyk-Timan theory) one obtains bounds for the error of approximation which depend on x and contain one of the quantities

$$\text{(1.1)} \qquad \begin{cases} \Delta_n := \Delta_n(x) := \sqrt{1-x^2}\,n^{-1} + n^{-2}\,, & n = 1,2,\dots\,, \quad \Delta_0 := 1 \\ \delta_n(x) := \sqrt{1-x^2}\,n^{-1} & , \ n = 1,2,\dots\,, \quad \delta_0 := 1\,. \end{cases}$$

One can then prove that for $f \in C(A)$, $r = 1,2,\dots$, there exist polynomials P_n for which

$$\text{(1.2)} \qquad |f(x) - P_n(x)| \le C_r\omega_r(f, \Delta_n(x))\,, \qquad n = 0,1,\dots\,.$$

Consequently for $f \in C^r$,

$$\text{(1.3)} \qquad |f(x) - P_n(x)| \le C_r\Delta_n(x)^r\omega(f^{(r)}, \Delta_n(x))\,.$$

Relations (1.3) have been proved by Dzyadyk and Timan (see their books: Dzyadyk [A-1977], Timan [A-1963]), and (1.2) by Dzyadyk and Freud for $r = 2$, by Brudnyi [1963] for any $r \ge 1$.

The first theorem of this type, by Nikolskii was slightly less favorable than the case $r = 0$ of (1.3). (See also Lemma 3.1.) Our proof of (1.3) in §4 and of (1.2) in §5 will be by means of a linear polynomial operators, in which case (1.2) and (1.3) are equivalent (see Theorem 4.3 of Chapter 7). This proof, as well as that of an inverse theorem of §6, will depend on an estimate of the derivatives of polynomials, given in §2.

The main protagonists of the second theory are Ditzian and Totik (see their book [A-1987]). The weighted modulus of smoothness, $\omega_r^\phi(f,t)_p$, (see §6 of Chapter 6), with $\phi(x) = \sqrt{1-x^2}$, $x \in [-1,1]$, plays an important role here. For example, we prove in §7 that

$$\text{(1.4)} \qquad \|f - P_n\|_p \le C_r\omega_r^\phi\Big(f, \frac{1}{n}\Big)_p\,, \qquad f \in L_p(A)$$

can be achieved. A classical example is the function $f(x) = (1-x^2)^\gamma$, $0 < \gamma < 1$, $x \in A$, for which (1.4) gives the correct order for $E_n(f) = \mathcal{O}(n^{-2\gamma})$, while the results of §5, Chapter 7 yield only the upper bound $Cn^{-\gamma}$. See also Löfström [1985].

Earlier, other definitions of moduli of smoothness and some direct and inverse theorems have been given by Ivanov [1980], Stens [1978] and others. A quite different approach appears in Butzer, Stens and Wehlen [1979].

Which of the two theories is preferable? Certainly, this question is meaningless. However, the second theory proved to be more adaptable, applicable also for the L_p-norm, $p \ne \infty$, not just for the uniform norm as in (1.2). The first theory has also been extended to $p \ne \infty$ by (Oswald [1978]), but his characterization is not simple.

§ 2. Local Inequalities for Polynomials

We shall derive some inequalities of local character on $A := [-1,1]$ which are of Bernstein type, and involve the quantity $\Delta_n(x) := \sqrt{1-x^2}\,n^{-1} + n^{-2}$. The subintervals $A_n := [-\alpha_n, \alpha_n]$, $\alpha_n := \sqrt{1-n^{-2}}$ are important for the asymptotic behavior of the $\Delta_n(x)$:

$$\text{(2.1)} \quad \Delta_n(x) \sim \sqrt{1-x^2}\,n^{-1} \ \text{ for } \ x \in A_n\,, \qquad \Delta_n(x) \sim n^{-2}\,, \ x \in A \setminus A_n\,;$$

(equivalently, one can take $\alpha_n' := \sqrt{1-cn^{-2}}$ for any fixed a constant $c > 0$). As an example of application of the set A_n we can mention: *If for $P_n \in \mathcal{P}_n$, one has $|P_n(x)| \le M$, $x \in A_n$, then for an absolute constant C,*

$$\text{(2.2)} \qquad |P_n(x)| \le CM\,, \qquad x \in A\,.$$

This follows by applying to the polynomials

$$\text{(2.3)} \qquad Q_n(y) := P_n(\alpha_n y)\,,$$

the inequality (2.9) of Chapter 4. Indeed, $|P_n(x)| \le M$, $x \in A_n$, means that $|Q_n(y)| \le M$, $y \in A$, and from (2.9) one computes $|Q_n(y)| \le CM$ if $|y| \le \alpha_n^{-1}$, which is (2.2).

As in §2 of Chapter 7, (but with slight changes of notation) we define the Jackson kernels for $m, n = 1,2,\dots$

$$\text{(2.4)} \qquad K_{n,m}(t) := \lambda_{n,m}\left(\frac{\sin\frac{nt}{2}}{n\sin\frac{t}{2}}\right)^{2m}, \qquad \int_{\mathbb{T}} K_{n,m}\,dt = 1\,;$$

they are even trigonometric polynomials of degree $m(n-1)$. One has (see Lemma 2.1 of Chapter 7), $\lambda_{n,m} \sim n$ and

$$\int_{-\pi}^{\pi} |t|^k K_{n,m}(t)\,dt \le C_m n^{-k}, \qquad k = 0, 1, \dots, 2m-2. \tag{2.5}$$

For the derivative T_n' of a trigonometric polynomial T_n we shall need a representation formula of Brudnyi [1959], which is a useful variation of a related formula of M. Riesz. For the purpose of the next theorem, we define $K_{n,0}(t) := \lambda_{n,0} := n$, $n = 1, 2, \dots$.

Theorem 2.1. *There exist trigonometric polynomials $H_{n,m}$ with the property*

$$|H_{n,m}| \le C_m n \tag{2.6}$$

so that for the kernel $\Lambda_{n,m}(t) := H_{n,m}(t)K_{n,m}(t)$ with $m = 0, 1, \dots$ one has

$$T_n'(t) = (T_n * \Lambda_{n,m})(t) = \frac{1}{\pi}\int_{\mathbb{T}} T_n(t-u)\Lambda_{n,m}(u)\,du. \tag{2.7}$$

Proof. For $m = 0$, we differentiate the representation $T_n = T_n * D_n$, where D_n is the Dirichlet kernel $D_n(u) = \frac{1}{2} + \cos u + \cdots + \cos nu$ and obtain

$$T_n' = T_n * D_n', \qquad D_n'(u) = -\sin u - \cdots - n \sin nu.$$

Thus, we can put $H_{n,0} := \frac{1}{n} D_n'$, $\Lambda_{n,0}(t) = D_n'$ and have $|H_{n,0}(u)| \le n$.

For $m > 0$ we prove our assertion by induction. If it is valid for some $m > 0$ and all $n = 0, 1, \dots$, then we have, for each trigonometric polynomial S_{2n} of degree $\le 2n$,

$$S_{2n}'(0) = \frac{1}{\pi}\int_{\mathbb{T}} S_{2n}(-u)\Lambda_{2n,m}(u)\,du.$$

We apply this to the polynomial of u

$$S_{2n}(u,t) := T_n(t+u)\left(\frac{\sin\frac{nu}{2}}{\sin\frac{u}{2}}\right)^2.$$

Then $S_{2n}'(0,t) = n^2 T_n'(t)$ and therefore

$$T_n'(t) = n^{-2}\frac{1}{\pi}\int_{\mathbb{T}} T_n(t-u)\left(\frac{\sin\frac{nu}{2}}{\sin\frac{u}{2}}\right)^2 \Lambda_{2n,m}(u)\,du.$$

Since $\lambda_{2n,m} \sim \lambda_{n,m+1}$, we obtain (2.6) and (2.7) if we select

$$H_{n,m+1}(u) := 2^{2m}\frac{\lambda_{2n,m}}{\lambda_{n,m+1}}\left(\cos\frac{nu}{2}\right)^{2m} H_{2n,m}(u). \qquad \square$$

Our inequalities will contain functions of the following class Ω_r, $r = 0, 1, \dots$. A continuous positive function $\Phi(u)$, $u > 0$, belongs to Ω_r if

$$\Phi(u) \le C(\lambda_0)\Phi(\lambda u) \quad \text{for all} \quad \lambda \ge \lambda_0 > 0 \tag{2.8}$$

and if Φ, for some absolute constant C, satisfies

$$\Phi(\lambda u) \le C\lambda^r \Phi(u), \qquad \lambda \ge 1. \tag{2.9}$$

Examples. **1.** The functions 1, u^β $(0 < \beta \le r)$, $\omega_1(u^r)$, $u^{r-1}\omega_1(u)$, $\omega_r(u)$ all belong to Ω_r if ω_1 and ω_r are some moduli of continuity or smoothness.

2. If $\Phi \in \Omega_r$, then $\Phi_1(u) := u^{-1}\Phi(u) \in \Omega_{r-1}$. Plainly, $\Omega_{r-1} \subset \Omega_r$. Inequality (2.9) implies

$$\Phi(u+v) \le C\big(\Phi(u) + \Phi(v)\big), \qquad u, v > 0, \tag{2.10}$$

and from (2.8), $\Phi(u) \le C\Phi(v)$ if $0 < u \le v$.

Lemma 2.2. *For $P_n \in \mathcal{P}_n$, $\Phi \in \Omega_r$ let*

$$|P_n(x)| \le M\Delta_n(x)^{-r}\Phi\big(\Delta_n(x)\big), \qquad x \in A_n. \tag{2.11}$$

Then this inequality, with M replaced by CM, $C = C(r)$, prevails also on A. The latter inequality follows also from

$$|P_n(x)| \le M\frac{n}{\sqrt{1-x^2}}\Delta_n(x)^{-r+1}\Phi\big(\Delta_n(x)\big), \qquad x \in A. \tag{2.12}$$

Proof. Since $\Delta_n(x) \ge n^{-2}$ and since Φ satisfies (2.9), relations (2.11) and (2.2) imply

$$|P_n(x)| \le CMn^{2r}\Phi(n^{-2}), \qquad x \in A. \tag{2.13}$$

For $x \in A \setminus A_n$, we have $n^{-2} \ge C\Delta_n(x)$ there, and because of (2.8), we have $|P_n(x)| \le CM\Delta_n(x)^{-r}\Phi(\Delta_n(x))$. We see that the required inequality holds for all $x \in A$. Furthermore, (2.12) implies (2.11). $\square$

The following inequality, due to Lebed [1957] and Brudnyi [1959], is the main result of this section. Special cases of it have been given by Dzyadyk and Timan.

Theorem 2.3. *There is a constant C_r with the following property. For each polynomial $P_n \in \mathcal{P}_n$ and each function $\Phi \in \Omega_r$, $r = 0, 1, \dots$, the inequality*

$$|P_n(x)| \le \Phi\big(\Delta_n(x)\big), \qquad x \in A = [-1, 1] \tag{2.14}$$

implies for $k = 0, \dots, r$,

$$|P_n^{(k)}(x)| \le C_r\Delta_n(x)^{-k}\Phi\big(\Delta_n(x)\big), \qquad x \in A \tag{2.15}$$

and

$$|P_n^{(r+1)}(x)| \le C_r n^{2(r+1)} \Phi(n^{-2}), \qquad x \in A. \tag{2.16}$$

The latter two inequalities also follow if (2.14) *holds only on* A_n.

Proof. If P_n satisfies (2.14), then for $T_n(t) := P_n(x)$, $x = \cos t$ we have

$$|T_n(t)| \le \Phi(\lambda_n(t)), \qquad \lambda_n(t) := |\sin t| n^{-1} + n^{-2}$$

and therefore by (2.10)

$$\begin{aligned}|T_n(t-u)| &\le C\Phi((|\sin t| + |u|)n^{-1} + n^{-2}) \\ &\le C[\Phi(\lambda_n(t)) + \Phi(|u|n^{-1} + n^{-2})].\end{aligned}$$

For the second term in the brackets,

$$\Phi(|u|n^{-1} + n^{-2}) \le C(|u|n+1)^r \Phi(n^{-2}) \le C(|u|^r n^r + 1)\Phi(n^{-2}).$$

Therefore, Theorem 2.1, (2.5) and (2.6) for $m = r+1$ yield

$$\begin{aligned}|T_n'(t)| &\le \frac{1}{\pi}\int_{\mathbb{T}} |T_n(t-u)|\,|\Lambda_{n,r+1}(u)|\,du \\ &\le Cn\Phi(\lambda_n(t)) + C\Phi(n^{-2})\int_{\mathbb{T}}(|u|^r n^r + 1)|\Lambda_{n,r+1}(u)|\,du \\ &\le Cn\Phi(\lambda_n(t)),\end{aligned} \tag{2.17}$$

that is,

$$|P_n'(x)| = |\sin t|^{-1}|T_n'(t)| \le \frac{Cn}{\sqrt{1-x^2}}\Phi(\Delta_n(x)), \qquad x \in A. \tag{2.18}$$

We can now prove the theorem by induction on r. If $r = 1$, the last inequality is identical with (2.12) for P_n'. We obtain (2.15) for $k = 1$ by means of Lemma 2.2. Then (2.16) follows from Markov's inequality.

Let now the theorem be known for $r-1$ instead of r, and let P_n satisfy (2.14) with $\Phi \in \Omega_r$. From (2.18),

$$|P_n'(x)| \le C\Delta_n(x)^{-1}\Phi(\Delta_n(x)) = C\Phi_1(\Delta_n(x)), \qquad x \in A_n,$$

where Φ_1 is the function $\Phi_1(u) := u^{-1}\Phi(u)$, which belongs to Ω_{r-1}. For the polynomials Q_n of (2.3) this means that $|Q_n'(y)| \le C\Phi_1(\Delta_n(y))$, $y \in A$. Indeed, if $x = \alpha_n y$, then

$$\begin{aligned}\Delta_n(y) \le \Delta_n(x) &= \frac{1}{n}\sqrt{1 - \Big(1 - \frac{1}{n^2}\Big)y^2} + \frac{1}{n^2} \\ &\le \frac{1}{n}\sqrt{1-y^2} + \frac{2}{n^2} \le 2\Delta_n(y),\end{aligned}$$

so that $\Delta_n(x) \sim \Delta_n(y)$. The inductive assumption yields for $k = 1, \dots, r$

$$|Q_n^{(k)}(y)| \le C_r \Delta_n(y)^{-k+1}\Phi_1(\Delta_n(y)), \qquad y \in A. \tag{2.19}$$

For P_n we get

$$|P_n^{(k)}(x)| \le C_r\Delta_n(x)^{-k}\Phi(\Delta_n(x)), \qquad x \in A_n. \tag{2.20}$$

On $A \setminus A_n$, $\Delta_n(x) \sim n^{-2}$ and we have only to show that $|P_n^{(k)}(x)| \le C_r n^{2k}\Phi(n^{-2})$ on this set. For $k = r$ this follows from Lemma 2.2. More generally, having verified this for the $(k+1)$-st derivative, we obtain for $\alpha_n < x \le 1$, that

$$|P_n^{(k)}(x)| \le |P_n^{(k)}(\alpha_n)| + \int_{\alpha_n}^x |P_n^{(k+1)}(u)|\,du \le Cn^{2k}\Phi(n^{-2}).$$

Moreover, (2.16) follows from (2.15), $k = r$, and Markov's inequality.

Suppose now that we know the validity of (2.14) only on the set A_n. We again consider the polynomials $Q_n(y) := P_n(x)$, $x = \alpha_n y$, of (2.3). Since $\Delta_n(y) \sim \Delta_n(x)$, $y \in A$, we have $|Q_n(y)| \le C\Phi(\Delta_n(y))$, $y \in A$. By what we have already proved, Q_n satisfies (2.15) and (2.16). It follows that (2.20) is valid for P_n and therefore the proof of (2.15) and (2.16) for P_n can be completed as above. □

§ 3. Properties of the Jackson Operators $P_{n,m}(f)$

By means of the kernel (2.4), which has the properties $\lambda_{n,m} \sim n$ and (2.5), we define the trigonometric Jackson integral $J_{n,m}(f,t) = \int_{\mathbb{T}} f(t+u)K_{n,m}(u)\,du$ and its algebraic version

$$P_{n,m}(f,x) = \int_{\mathbb{T}} f(\cos(t+u))K_{n,m}(u)\,du, \tag{3.1}$$

where $f \in C(A)$, $A := [-1,1]$, $x = \cos t$, $0 \le t \le \pi$. The $J_{n,m}$ are even trigonometric polynomials of degree $\le m(n-1)$, consequently the *Jackson polynomials* $P_{n,m}$ are algebraic polynomials in $\mathcal{P}_{m(n-1)}$.

The simplest case of (1.3), when $r = 1$, can be handled by the operator $P_{n,2}$:

Lemma 3.1. *If* $\omega(t) = \omega(f,t)$ *is the modulus of continuity of* $f \in C(A)$, *then*

$$|f(x) - P_{n,2}(f,x)| \le C\omega(\Delta_n(x)), \qquad x \in A. \tag{3.2}$$

Proof. With $x = \cos t$ we have

$$\begin{aligned}|f(x) - P_{n,2}(f,x)| &= \left|\int_{\mathbb{T}}\big[f(\cos t) - f(\cos(t+u))\big]K_{n,2}(u)\,du\right| \\ &\le \int_{\mathbb{T}} \omega\left(2\left|\sin\Big(t + \frac{u}{2}\Big)\sin\frac{u}{2}\right|\right)K_{n,2}(u)\,du.\end{aligned}$$

For $-\pi \le u, t \le \pi$, the argument of ω does not exceed

$$|u|\left|\sin\left(t+\frac{u}{2}\right)\right| \le |u|(|\sin t|+|u|) \le \left(\frac{|\sin t|}{n}+\frac{1}{n^2}\right)(n|u|+n^2u^2)$$
$$= (n|u|+n^2u^2)\Delta_n(x)\,.$$

Since $\omega(a\delta) \le (a+1)\omega(\delta)$, by (2.5),

$$|f(x)-P_{n,2}(f,x)| \le C\omega\big(\Delta_n(x)\big)\int_{\mathbb{T}}(n|u|+n^2u^2+1)K_{n,2}(u)\,du \le C\omega\big(\Delta_n(x)\big)\,.$$

□

The general case of (1.3) will be proved by induction, and for this purpose we employ

Lemma 3.2. *Let $f \in C^1(A)$, $A = [-1,1]$, $\Phi \in \Omega_r$. If*

(3.3) $$|f'(x)| \le \Delta_n(x)^{-1}\Phi\big(\Delta_n(x)\big)\,, \qquad x \in A\,,$$

then

(3.4) $$|f(x)-P_{n,r+1}(f,x)| \le C_r\Phi\big(\Delta_n(x)\big)\,,$$

and

(3.5) $$|P^{(k)}_{n,r+1}(f,x)| \le C_r\Delta_n(x)^{-k}\Phi\big(\Delta_n(x)\big)\,, \qquad k=1,\dots,r\,.$$

Proof. From (3.1), if $x = \cos t$,

(3.6) $$f(x)-P_{n,r+1}(f,x) = \int_{\mathbb{T}}\Big[f(\cos t)-f\big(\cos(t+s)\big)\Big]K_{n,r+1}(s)\,ds\,.$$

Since $\frac{d}{ds}f(\cos(t+s)) = -f'(\cos(t+s))\sin(t+s)$,

$$f(\cos t)-f\big(\cos(t+s)\big) = \int_0^s f'\big(\cos(t+u)\big)\sin(t+u)\,du\,,$$

and for the difference (3.6) we obtain, changing the order of integration,

(3.7) $$\int_0^\pi\Big[f'\big(\cos(u+t)\big)\sin(u+t)+f'\big(\cos(u-t)\big)\sin(u-t)\Big] \times \int_u^\pi K_{n,r+1}(s)\,ds\,du\,.$$

For $0 \le u,t \le \pi$, if $y = \cos(u+t)$, we have

$$\Delta_n(y) \le \frac{1}{n}(u+\sin t)+\frac{1}{n^2} = \frac{1}{n}\big(\sqrt{1-x^2}+u\big)+\frac{1}{n^2} \le (nu+1)\Delta_n(x)\,.$$

Thus, by (3.3) and (2.9),

(3.8) $$\begin{aligned}|f'(y)\sin(u+t)| &\le Cn\Phi\big((nu+1)\Delta_n(x)\big)\\ &\le Cn(nu+1)^r\Phi\big(\Delta_n(x)\big)\,,\end{aligned}$$

and similarly if t is replaced by $-t$. Using (3.8) in (3.7) and integrating by parts, we obtain that (3.7) does not exceed

$$\begin{aligned}&C_r\Phi\big(\Delta_n(x)\big)\int_0^\pi n(nu+1)^r\int_u^\pi K_{n,r+1}(s)\,ds\,du\\ &= C_r\Phi\big(\Delta_n(x)\big)\left[\frac{-1}{r+1}\int_0^\pi K_{n,r+1}(s)\,ds+\int_0^\pi\frac{(nu+1)^{r+1}}{(r+1)}K_{n,r+1}(u)\,du\right]\\ &\le C_r\Phi\big(\Delta_n(x)\big)\,.\end{aligned}$$

This establishes (3.4). Now, differentiating $P_{n,r+1}(f,x)$, we get

$$P'_{n,r+1}(f,x) = \frac{1}{\sqrt{1-x^2}} \times \int_0^\pi\Big[f'\big(\cos(u+t)\big)\sin(u+t)-f'\big(\cos(u-t)\big)\sin(u-t)\Big]K_{n,r+1}(u)\,du\,.$$

The estimation of this integral is similar to that of (3.7), but simpler. By means of (3.8) and (2.5) we get

(3.9) $$|P'_{n,r+1}(f,x)| \le C_r\frac{n}{\sqrt{1-x^2}}\Phi\big(\Delta_n(x)\big)\,.$$

Now the polynomials $P'_{n,r+1}$ belong to the class $\mathcal{P}_{n_1}$, $n_1 := (r+1)n$, and $(r+1)^{-2}\Delta_n(x) \le \Delta_{n_1}(x) \le \Delta_n(x)$. It follows that

$$|P'_{n,r+1}(x)| \le C_r\Phi_1\big(\Delta_{n_1}(x)\big)\,, \qquad x \in A_n\,,$$

with $\Phi_1(t) := t^{-1}\Phi(t)$. Now $\Phi_1 \in \Omega_{r-1}$. Therefore applying the second part of Theorem 2.3 we see that the last inequality persists on A. Another application of Theorem 2.3 then yields (3.5). □

§ 4. Simultaneous Approximation of Functions and Their Derivatives

We are now able to prove the relation (1.3) of Timan [1951]. Actually, we shall establish more, namely the possibility of *simultaneous approximation* (Trigub [1962]), when the function f and its derivatives $f^{(k)}$, $k \le r$, are approximated by a single polynomial P_n and its derivatives. Simple proofs have been provided by Malozemov, [A-1973].

Theorem 4.1. *For each $r = 0,1,\dots$ there exists an integral polynomial operator $Q_{n,r}(f)$ of degree $(r+2)n$ and a constant $C_r > 0$ with the property that for any function $f \in C^r(A)$, and for $k = 0,\dots,r$, one has*

(4.1) $$|f^{(k)}(x)-Q^{(k)}_{n,r}(f,x)| \le C_r\Delta_n(x)^{r-k}\omega\big(f^{(r)},\Delta_n(x)\big)\,, \qquad x \in A\,.$$

We shall occasionally abbreviate $\Delta_n := \Delta_n(x)$, $\omega_n := \omega(f^{(r)},\Delta_n(x))$.

The idea of the proof is as follows. We approximate $f^{(r)}$ by the polynomial $P_{n,2}(f^{(r)})$ of the previous section. By integration, this gives an approximation to $f^{(r-1)}$ which is not quite good enough. We make a correction to this approximation by using the operator $P_{n,3}$ and continue in this manner, integrating and making corrections. To facilitate the description of this construction, we introduce the following notation. Let $P_{n,m}$ be the polynomial operators (3.1) (they are of degree $\le m(n-1)$), let I denote the identity operator, and let $J(f,x) := \int_{-1}^{x} f(u)\,du$. For $f \in C^r(A)$, we put

$$Q_{n,r}(f) := f - (I - P_{n,r+2})\big[J(I-P_{n,r+1})\big]\dots\big[J(I-P_{n,2})\big]f^{(r)}\,. \tag{4.2}$$

We can express this as a sum of the functions $f - J^r f^{(r)}$ and $P_{n,m}g$ where g are functions given by

$$g_0 = f^{(r)}\,,\quad g_m = J(I-P_{n,m+1})(g_{m-1})\,,\qquad m = 1,\dots,r\,. \tag{4.3}$$

We have

$$\begin{aligned} f - Q_{n,r}(f) &= (I-P_{n,r+2})(g_r) = -P_{n,r+2}(g_r) + J(I-P_{n,r+1})(g_{r-1})\\ &= -P_{n,r+2}(g_r) - JP_{n,r+1}(g_{r-1}) + J^2(I-P_{n,r})(g_{r-2}) = \cdots\\ &= -\sum_{m=0}^{r}(J^{r-m}P_{n,m+2})(g_m) + J^r(g_0)\,. \end{aligned} \tag{4.4}$$

The m-th term of the last sum is a polynomial of degree $\le (m+2)(n-1)+r-m$, while $f - J^r(g_0) = f - J^r(f^{(r)})$ is a polynomial of degree $\le r-1$. This establishes that *$Q_{n,r}(f)$ is a polynomial of degree $< (r+2)n$.*

Proof of Theorem 4.1. We prove a more general relation than (4.1). Namely, for $s = 0,\dots r$, $k = 0,\dots s$, we have

$$|g_s^{(k)}(x) - P_{n,s+2}^{(k)}(g_s,x)| \le C_r \Delta_n(x)^{s-k}\omega_n\,. \tag{4.5}$$

For $s = r$, this is the relation (4.1) to be proved, because $g_r - P_{n,r+2}(g_r) = f - Q_{n,r}(f)$. For $s = 0$, $k = 0$, (4.5) is known: it follows from Lemma 3.1 applied to $f^{(r)}$. We shall prove (4.5) by induction on s. We assume that (4.5) holds with s replaced by $s-1$ and all $k \le s-1$. We thus have

$$\left|\frac{d^k}{dx^k}(I-P_{n,s+1})(g_{s-1},x)\right| \le C_r\Delta_n^{s-1-k}\omega_n\,,\quad k = 0,\dots,s-1,\ \ x\in A\,. \tag{4.6}$$

Since $g_s' = \frac{d}{dx}[J(I-P_{n,s+1})](g_{s-1}) = (I-P_{n,s+1})(g_{s-1})$, from (4.6) with $k = 0$ we obtain $|g_s'(x)| \le C_r\Delta_n^{s-1}\omega_n$, $x \in A$. We can apply Lemma 3.2 with $\Phi(u) = u^s\omega(f^{(r)},u)$ which belongs to the class Ω_s. In the first place, $|g_s - P_{n,s+2}(g_s)| \le C_r\Delta_n^s\omega_n$, this is (4.5) for $k = 0$. Next, for $k = 1,\dots,s$,

$$|P_{n,s+2}^{(k)}(g_s,x)| \le C_r\Delta_n^{s-k}\omega_n\,. \tag{4.7}$$

Now (4.5) for $k = 1,\dots,s$ follows from the identity

$$\frac{d^k}{dx^k}(I-P_{n,s+2})(g_s,x) = -P_{n,s+2}^{(k)}(g_s,x) + \frac{d^{k-1}}{dx^{k-1}}(I-P_{n,s+1})(g_{s-1},x)\,.\qquad\square$$

Remark. It should be noted that (1.3) is true even with $\omega_2(f^{(r)},t)$ instead of $\omega(f^{(r)},t)$, see Dzyadyk [1958] and [A-1977].

Sometimes, (see the proof of Theorem 4.8), one needs an addition to Theorem 4.1, given below as Theorem 4.3.

Lemma 4.2. *If $f \in C(A)$, then for all $x \in A$,*

$$|P_{n,2}'(f,x)| \le C\delta_n(x)^{-1}\omega(f,\delta_n(x))\,,\qquad \delta_n(x) := \frac{\sqrt{1-x^2}}{n}\,. \tag{4.8}$$

Proof. Differentiating the polynomial $P_{n,2}$ of (3.1), we get with $x = \cos t$

$$P_{n,2}'(f,x) = \frac{1}{\sqrt{1-x^2}}\int_{\mathbb{T}} f(\cos u)K_{n,2}'(u-t)\,du\,,$$

and since $K_{n,2}'(u)$ is odd,

$$P_{n,2}'(f,x) = \frac{1}{\sqrt{1-x^2}}\int_0^{\pi}\Big[f\big(\cos(u+t)\big) - f\big(\cos(u-t)\big)\Big]K_{n,2}'(u)\,du\,.$$

This yields

$$\begin{aligned} |P_{n,2}'(f,x)| &\le \frac{1}{\sqrt{1-x^2}}\int_0^{\pi}\omega(2|\sin t\sin u|)|K_{n,2}'(u)|\,du\\ &\le \frac{1}{\sqrt{1-x^2}}\omega\left(\frac{\sqrt{1-x^2}}{n}\right)4\lambda_{n,2}\int_0^{\pi}(2nu+1)L_n(u)\,du\,,\\ L_n(t) &= \left|\left(\frac{\sin\frac{nt}{2}}{n\sin\frac{t}{2}}\right)^3\frac{d}{dt}\left(\frac{\sin\frac{nt}{2}}{n\sin\frac{t}{2}}\right)\right|\,. \end{aligned} \tag{4.9}$$

It is easy to prove that the absolute value of the last derivative is $\le n$. Indeed, if n is even, $n = 2k$, we can use the formula

$$\cos\frac{t}{2} + \cdots + \cos\frac{2k-1}{2}t = \frac{\sin kt}{2\sin\frac{t}{2}}\,,$$

which yields

$$\begin{aligned} \left|\frac{d}{dt}\left(\frac{\sin kt}{n\sin\frac{t}{2}}\right)\right| &= \frac{2}{n}\left|\frac{d}{dt}\left(\cos\frac{t}{2} + \cdots + \cos\frac{2k-1}{2}t\right)\right|\\ &\le \frac{2}{n}\left(\frac{1}{2} + \cdots + \frac{2k-1}{2}\right) \le n\,. \end{aligned}$$

If n is odd, $n = 2k+1$, we use in a similar manner a formula pertaining to the Dirichlet kernel. Since $\lambda_{n,1} \sim n$, we get

$$|P'_{n,1}(f,x)| \le C\delta_n(x)^{-1}\omega\big(\delta_n(x)\big) \int_0^\pi n(2nu+1)\left|\frac{\sin\frac{nu}{2}}{n\sin\frac{u}{2}}\right|^3 du$$

and the integral does not exceed $C\int_0^\infty (4t+1)|\frac{\sin t}{t}|^3\,dt < +\infty$. □

We can now derive

Theorem 4.3 (Trigub). *The polynomial operator $Q_{n,r}(f)$ of Theorem 4.1 satisfies*

$$(4.10)\qquad |Q_{n,r}^{(r+1)}(f,x)| \le C_r n^2 \omega(f^{(r)}, n^{-2}).$$

Proof. Differentiating $Q_{n,r}$ in (4.4), we obtain, since $(f - J^r(g_0))^{(r+1)} = 0$,

$$(4.11)\qquad Q_{n,r}^{(r+1)}(f) = \sum_{m=0}^{r} P_{n,m+2}^{(m+1)}(g_m).$$

To estimate the terms $m = 1, \dots, r$ of this sum, we note that for $k = s = m$ the relation (4.7) yields

$$|P_{n,m+2}^{(m)}(g_m,x)| \le C_r\omega\big(f^{(r)}, \Delta_n(x)\big),$$

and then from Theorem 2.3 we have

$$|P_{n,m+2}^{(m+1)}(g_m,x)| \le C'_r n^2\omega(f^{(r)}, n^{-2}), \qquad x \in A.$$

For the term $m = 0$ of the sum we apply (4.8):

$$|P'_{n,2}(g_0,x)| = |P'_{n,2}(f^{(r)},x)| \le C\delta_n(x)^{-1}\omega\big(f^{(r)}, \delta_n(x)\big).$$

A modulus of continuity ω satisfies $\frac{\omega(u_1)}{u_1} \le 2\frac{\omega(u)}{u}$ if $0 < u \le u_1$. For $x \in A_n$, $\delta_n(x) \ge n^{-2}$, and we get

$$(4.12)\qquad |P'_{n,2}(f^{(r)},x)| \le Cn^2\omega(f^{(r)}, n^{-2}), \qquad x \in A_n.$$

By means of (2.2), we extend (4.12) onto A, completing the proof. □

A companion result to (4.1) are the estimates (4.17) of Leviatan [1982]. They deal with polynomials of best approximation and contain $E_n(f^{(k)})$ instead of $\omega(f^{(r)}, \Delta_n)$.

Lemma 4.4. *Let $F \in C^1$, $|F'(x)| \le M$ on A. Then for the polynomials $R_n^* \in \mathcal{P}_n$ of best uniform approximation to F,*

$$(4.13)\qquad |R_n^{*(k)}(x)| \le C_r M\frac{1}{n\Delta_n(x)^k}, \qquad k = 1, 2, \dots.$$

Proof. According to Theorem 6.1 of Chapter 7, for some $P_n \in \mathcal{P}_n$,

$$(4.14)\qquad \|F - P_n\| \le CM\frac{1}{n}, \quad \|P'_n\| \le CM.$$

From the first relation, $\|P_n - R_n^*\| \le CM\frac{1}{n}$ and by Theorem 2.3 (since constants belong to Ω_r for each r),

$$(4.15)\qquad |P_n^{(k)}(x) - R_n^{*(k)}(x)| \le \frac{CM}{n\Delta_n(x)^k}.$$

Similarly, from the second relation in (4.14),

$$(4.16)\qquad |P_n^{(k)}(x)| \le \frac{CM}{\Delta_n(x)^{k-1}} \le \frac{CM}{n\Delta_n(x)^k}.$$

The two inequalities imply (4.13) □

Theorem 4.5. *If $f \in C^r(A)$, then for its polynomial P_n^* of best approximation*

$$(4.17)\qquad |f^{(k)}(x) - P_n^{*(k)}(x)| \le C_r\frac{1}{(n\Delta_n(x))^k}E_{n-k}(f^{(k)}), \quad k = 0, \dots, r, \quad x \in A.$$

Proof. For $r = 0$, $k = 0$ this is obvious. Suppose (4.17) is true for some r. Let $f \in C^{r+1}(A)$ and let $Q'_n \in \mathcal{P}_{n-1}$ be the best approximation to f'. If Q_n is an integral of Q'_n, then for $k = 0, \dots, r$, $x \in A$

$$\begin{aligned} |f^{(k+1)}(x) - Q_n^{(k+1)}(x)| &\le C_r\frac{1}{(n\Delta_n)^k}E_{n-k}(f^{(k+1)}) \\ &\le C_r\frac{1}{(n\Delta_n)^{k+1}}E_{n-k-1}(f^{(k+1)}). \end{aligned}$$

We put $F := f - Q_n$. This function satisfies $\|F'\| = E_{n-1}(f')$. For the polynomials R_n^* of best approximation to F, from (4.13), and the inequality $E_{n-1}(f') \le n^{-k}E_{n-k-1}(f^{(k+1)})$, we get

$$|R_n^{*(k+1)}(x)| \le C_r\frac{E_{n-1}(f')}{n\Delta_n^{k+1}} \le C_r\frac{E_{n-k-1}(f^{(k+1)})}{(n\Delta_n)^{k+1}}, \qquad x \in A.$$

Obviously, $P_n^* = Q_n + R_n^*$. Combining the two last inequalities, we obtain (4.17) for $k = 1, \dots, r+1$ and the case $k = 0$ is trivial. □

Since $\frac{1}{n\Delta_n} \le n$, $x \in A$, we can derive from (4.17) estimates of the error which do not depend on x. They should be compared with (6.4) of Chapter 7.

Corollary 4.6. *For $f \in C^r[-1,1]$, $0 \le k \le r$, the polynomials P_n^* of best approximation to f satisfy*

$$\|f^{(k)} - P_n^{*(k)}\| \le C_r n^k E_{n-k}(f^{(k)}) \le C_r n^{2k-r} E_{n-r}(f^{(r)}), \quad k = 0, \dots, r. \tag{4.18}$$

The history of the next Theorem 4.8 is interesting. One of the authors (Lorentz [1964]) raised the question whether $\Delta_n(x)$ can be replaced by the smaller quantity $\delta_n(x) := \sqrt{1-x^2}/n$ in polynomial approximation on $[-1,1]$; Telyakovskii [1966] gave a positive answer, proving Theorem 4.8 (that is, (4.19)) for $r=1$. Then Gopengauz [1967] gave the relation, $r = 1, 2, \dots$

$$|f(x) - P_n(x)| \le C_r \omega_r(f, \delta_n(x)), \qquad f \in C(A), \ A = [-1,1], \tag{4.19}$$

But his proof was incorrect. However, for $r=2$, DeVore gave a correct proof ([1976] and Theorem 5.4 below). Our proof of Theorem 4.8 is by Malozemov [A-1973] and v. Golitschek.

We shall need a lemma:

Lemma 4.7 (v. Golitschek). *For each $r = 1, 2, \dots$ and $n \ge 3r+2$ there exist $r+1$ polynomials U_k, $k = 0, \dots, r$, of degree $\le n$ which satisfy*

$$\begin{cases} U_k^{(k)}(0) = 1, & U_k^{(j)}(0) = 0, \quad j \ne k, \ 0 \le j \le r \\ U_k^{(j)}(1) = 0, & 0 \le j \le r, \\ |U_k(x)| \le K_r n^{-2k}, & 0 \le x \le 1 \end{cases} \tag{4.20}$$

where K_r is a constant depending on r.

Proof. We construct the U_k as products of two polynomials $Q_r \in \mathcal{P}_{2r+1}$ and $P_n \in \mathcal{P}_{n-2r-1}$. The first of them is independent of n and k, and is the solution of the interpolation problem

$$\begin{cases} Q_r(0) = 1, & Q_r^{(j)}(0) = 0, \quad j = 1, \dots, r \\ Q_r^{(j)}(1) = 0, & j = 0, \dots, r. \end{cases}$$

The second factor $P = P_{r,k,n}$ is the unique polynomial of the form

$$P(x) = x^k + \sum_{j=r+1}^{m} a_j x^j, \qquad m := n - 2r - 1$$

with the minimal uniform norm on $[0,1]$. According to Theorem 5.5 of Chapter 11,

$$\|P\|_\infty[0,1] \le \prod_{j=r+1}^{m} \frac{j-k}{j+k}.$$

Cancelling all like terms in the numerator and denominator of this product, we obtain

$$\|P\| \le \prod_{j=r+1}^{r+2k} (j-k) \Big/ \prod_{j=m-2k+1}^{m} (j+k) \le \left(\frac{2r}{m-r+1}\right)^{2k} \le K_r' n^{-2k}.$$

The polynomials $U_k := \frac{1}{k!} Q_r P$ satisfy (4.20) with $K_r = K_r' \|Q_r\|[0,1]$. □

Theorem 4.8 (Gopengauz). *Each function $f \in C^r(A)$, $r = 0, 1, \dots$, admits simultaneous approximation by a sequence of polynomials $P_n \in \mathcal{P}_n$ so that*

$$|f^{(k)}(x) - P_n^{(k)}(x)| \le C_r \delta_n(x)^{r-k} \omega(f^{(r)}, \delta_n(x)), \qquad k = 0, \dots, r, \ x \in A. \tag{4.21}$$

Proof. We use the polynomials $Q_{n,r}(f)$ of Theorems 4.1 and 4.3, and correcting them, obtain the P_n. The $Q_{n,r}(f)$ have degree $\le (r+2)n$. By putting $Q_n := Q_{[n/(r+2)],r}(f)$, we obtain, for a given $f \in C^r(A)$, a sequence of polynomials $Q_n \in \mathcal{P}_n$ with the properties for $x \in A$,

$$|f^{(k)}(x) - Q_n^{(k)}(x)| \le C_r \Delta_n(x)^{r-k} \omega(f^{(r)}, \Delta_n(x)), \qquad k = 0, \dots, r, \tag{4.22}$$

$$|Q_n^{(r+1)}(x)| \le C_r n^2 \omega(f^{(r)}, n^{-2}). \tag{4.23}$$

For $x_1 = -1$, $x_2 = 1$ we define

$$c_{k,j} := f^{(k)}(x_j) - Q_n^{(k)}(x_j), \qquad k = 0, \dots, r, \ j = 1, 2.$$

Since $\Delta_n(\pm 1) = n^{-2}$, from (4.22) we derive $|c_{k,j}| \le C_r n^{-2r+2k} \omega(f^{(r)}, n^{-2})$. We define the polynomials $S_n \in \mathcal{P}_n$ by means of

$$S_n(x) := \sum_{k=0}^{r} \left\{ c_{k,1} 2^k U_k\left(\frac{1+x}{2}\right) + c_{k,2}(-2)^k U_K\left(\frac{1-x}{2}\right) \right\}. \tag{4.24}$$

We claim that the polynomials $P_n := Q_n + S_n$ satisfy (4.21). From the definition of the S_n it follows that

$$f^{(k)}(\pm 1) - P_n^{(k)}(\pm 1) = 0, \qquad k = 0, \dots, r, \tag{4.25}$$

and from the estimates of the $c_{k,j}$

$$\|S_n\| \le \sum_{k=0}^{r} (|c_{k,1}| + |c_{k,2}|) 2^k K_r n^{-2k} \le C n^{-2r} \omega(f^{(r)}, n^{-2}).$$

Hence by Markov's inequality

$$\|S_n^{(k)}\| \le C n^{2k-2r} \omega(f^{(r)}, n^{-2}), \qquad k = 0, \dots, r+1. \tag{4.26}$$

To derive (4.21), we consider two cases. If $x \in A_n$, then $\Delta_n(x) \le 2\delta_n(x)$, and therefore by (4.22)

$$|f^{(k)}(x) - Q_n^{(k)}(x)| \le C \delta_n(x)^{r-k} \omega(f^{(r)}, \delta_n(x)).$$

Moreover, then $n^{-2} \le \delta_n(x)$, hence by (4.26) also $|S_n^{(k)}(x)|$ has this upper bound. This yields (4.21) for $x \in A_n$.

Let now $x \in A \setminus A_n$. We put $F := f - P_n$, and can restrict ourselves to the case $x > 0$. Then $1 - x \le \delta_n(x)$, moreover

$$(1-x)\omega(f^{(r)},n^{-2}) \le (1-x)\left(1+\frac{1}{n\sqrt{1-x^2}}\right)\omega(f^{(r)},\delta_n(x)) \le 2\delta_n(x)\omega\big(f^{(r)},\delta_n(x)\big)\,. \tag{4.27}$$

Because of (4.25), for $k=0,\dots,r-1$, the Taylor polynomial of $F^{(k)}$ at $x=1$ of degree $r-k$ is identically zero. Hence using the remainder formula for Taylor interpolation ((5.6) of Chapter 2), (4.26) and (4.27), we have for some point $y \in A\setminus A_n$,

$$\begin{aligned}|F^{(k)}(x)| &= (1-x)^{r-k}|F^{(r)}(y)|/(r-k)!\\ &\le (1-x)^{r-k}\Big\{|f^{(r)}(y)-Q_n^{(r)}(y)|+|S_n^{(r)}(y)|\Big\}\\ &\le C(1-x)^{r-k}\omega(f^{(r)},n^{-2}) \le C\delta_n(x)^{r-k}\omega\big(f^{(r)},\delta_n(x)\big)\,.\end{aligned} \tag{4.28}$$

On the other hand, if $k=r$, then for some $y\in(x,1)$, by (4.23), (4.26), and (4.27),

$$\begin{aligned}|F^{(r)}(x)| &= |F^{(r)}(x)-F^{(r)}(1)|\\ &\le |f^{(r)}(x)-f^{(r)}(1)|+|P_n^{(r)}(x)-P_n^{(r)}(1)|\\ &\le \omega(f^{(r)},1-x)+(1-x)|P_n^{(r+1)}(y)|\\ &\le \omega\big(f^{(r)},\delta_n(x)\big)+(1-x)\Big\{|Q_n^{(r+1)}(y)|+|S_n^{(r+1)}(y)|\Big\}\\ &\le \omega\big(f^{(r)},\delta_n(x)\big)+(1-x)n^2\omega(f^{(r)},n^{-2}) \le C\omega\big(f^{(r)},\delta_n(x)\big)\,.\end{aligned}$$

This, together with (4.28), establishes the theorem. □

§ 5. Brudnyi's Theorem

For arbitrary $r=1,2,\dots$, the possibility of approximation

$$|f(x)-P_n(x)| \le C_r\omega_r\big(f,\Delta_n(x)\big)\,, \qquad f\in C(A)\,, \quad x\in A[-1,1] \tag{5.1}$$

is much more delicate than that of (1.3). It has been established only in 1968 by Brudnyi [1968]. The special case $r=2$ has been treated by Dzyadyk [1958] and somewhat later by Freud [1959]. This follows neither from Theorem 5.1 of Chapter 7, for $\Delta_n(x)$ is not a constant, nor from the arguments of the previous section, which require f to be differentiable. The proof requires new approximation operators, among them the operators $S_{n,m}(g,t)$, $g\in C(\mathbb{T})$ of (2.8), Chapter 7:

$$S_{n,m}(g,t) = \int_{\mathbb{T}}\big[(-1)^{m+1}\Delta_u^m(g,t)+g(t)\big]K_{n,m}(u)\,du \tag{5.2}$$

where the $K_{n,m}$ are the Jackson kernels (2.4). We follow here the proof scheme of DeVore [1976]. We begin with a lemma about functions on the circle $\mathbb{T}$.

Lemma 5.1. *If a function $g\in C^k(\mathbb{T})$ has, for some $j,k\ge 0$ with $k\le r$, $k+j\le 2r$ the property*

$$|g^{(k)}(t)| \le M|\sin t|^j\,, \qquad t\in\mathbb{T}\,,$$

then

$$|S_n(g,t)-g(t)| \le CMn^{-k}(|\sin t|^j+n^{-j})\,, \qquad t\in\mathbb{T}\,. \tag{5.3}$$

Here $S_n := S_{n,2r}$ and the constant C depends on r.

Proof. We shall use the obvious inequality

$$(a+b)^k \le 2^k(a^k+b^k)\,, \qquad a,b\ge 0\,.$$

For $t,u\in\mathbb{T}$ there exists a ξ between 0 and ku for which (see §7 of Chapter 4)

$$\begin{aligned}|\Delta_u^k(g,t)| &= |g^{(k)}(t+\xi)|\;|u|^k \le M|\sin(t+\xi)|^j|u|^k\\ &\le CM(|\sin t|^j+|u|^j)|u|^k\,.\end{aligned}$$

With a maximum taken for all s between t and $t+(2r-k)u$,

$$\begin{aligned}|\Delta_u^{2r}(g,t)| &= |\Delta_u^{2r-k}\Delta_u^k(g,t)| \le 2^{2r-k}\max_s|\Delta_u^k(g,s)|\\ &\le CM|u|^k\max_s(|\sin s|^j+|u|^j)\\ &\le CM|u|^k(|\sin t|^j+|u|^j)\,.\end{aligned}$$

From the definition of the operators S_n and (2.5), we obtain

$$\begin{aligned}|S_n(g,t)-g(t)| &\le \int_{\mathbb{T}}|\Delta_u^{2r}(g,t)|K_{n,2r}(u)\,du\\ &\le CMn^{-k}\big[|\sin t|^j+n^{-j}\big]\,.\end{aligned}$$ □

Functions $f_0\in C^r(A)$ produce functions $g_0\in C^r(\mathbb{T})$ by means of the formula $g_0(t) := f_0(\cos t)$. The following lemma describes the structure of the functions g_0.

Lemma 5.2. *If $f_0\in C^r(A)$, and $M := \max_{1\le k\le r}\|f_0^{(k)}\|$, then g_0 has the representation*

$$g_0 = h_0+h_1+\cdots+h_r\,, \tag{5.4}$$

where $h_k\in C^{r+k}(\mathbb{T})$ and with a constant C depending only on r,

$$|h_k^{(r+k)}(t)| \le CM|\sin t|^{r-k}\,, \qquad t\in\mathbb{T}\,. \tag{5.5}$$

Proof. In the following formulas, we need sums of the type

$$R_{p,q}(t) := \sum_{q\le k<p} \tau_k(t)(\sin t)^{(p-2k)_+} f_0^{(p-k)}(\cos t) \tag{5.6}$$

where $0 \le q \le p \le 2r$ and the τ's, here and later, are some trigonometric polynomials, not depending on f_0, and $(p-2k)_+$ is the positive part of $p-2k$.

Since $f_0 \in C^r(A)$, the sums $R_{p,q}$ are defined and belong to $C(\mathbb{T})$ whenever $q \ge p-r$. Moreover, if $q > p-r$, we have even $R_{p,q} \in C^1(\mathbb{T})$. The derivative of a sum $R_{p,q}$ is a sum of type $R_{p+1,q}$. In particular, $g_0^{(p)}$ is a sum $R_{p,0}$ (with $\tau_0 = (-1)^p$) for $p = 0, 1, \dots, r$. Hence

$$g_0^{(r)}(t) = (-1)^r(\sin t)^r f_0^{(r)}(\cos t) + R_{r,1}\,. \tag{5.7}$$

The first term in (5.7) on the right has mean value zero. Let h_0 be its r^{th} periodic integral. Obviously, $h_0 \in C^r(\mathbb{T})$ and $|h_0^{(r)}| \le M|\sin t|^r$, which is (5.5) for $k = 0$.

Next, $g_1 := g_0 - h_0$ has r^{th} derivative $R_{r,1}$ and $r+1^{st}$ derivative $R_{r+1,1}$, so that

$$g_1^{(r+1)}(t) = \tau(t)(\sin t)^{(r-1)} f_0^{(r)}(\cos t) + R_{r+1,2}\,.$$

We define h_1 to be the $r+1^{st}$ periodic integral of the first term on the right. As required,

$$h_1 \in C^{r+1} \quad \text{and} \quad |h_1^{(r+1)}(t)| \le CM|\sin t|^{r-1}\,.$$

Continuing in this way we get

$$(g_0 - h_0 - \cdots - h_{r-1})^{(2r)} = R_{2r,r} = \sum_{r\le k<2r} \tau_k^*(t) f_0^{(2r-k)}(\cos t)\,.$$

If $h_r := g_0 - h_0 - \cdots - h_{r-1}$, then $h_r \in C^{2r}$ and

$$|h_r^{(2r)}(t)| \le C \sum_{1\le k\le r} \|f_0^{(k)}\|_\infty \le CM\,. \qquad \square$$

Theorem 5.3 (Brudnyi). *For each $r = 1, 2, \dots$, there is a sequence of linear algebraic polynomial operators S_n^* of degrees $\le n$ with the property that for $f \in C(A)$*

$$|f(x) - S_n^*(f,x)| \le C_r\omega_r(f, \Delta_n(x))\,, \qquad x \in A\,. \tag{5.8}$$

Proof. We construct the operators S_n^* on $C(A)$ as follows. To $f \in C(A)$, we assign $f_0 := f - Q_r f$, where $Q_r f$ is the polynomial of degree $< r$ which interpolates f at some r fixed distinct points of A. To f_0 corresponds the continuous function $g_0(t) := f_0(\cos t)$. To g_0 we apply the operator S_n of Lemma 5.1. Since $S_n(g_0, t)$ is an even trigonometric polynomial of degree $\le n$, after substitution $x = \cos t$ it becomes an algebraic polynomial $\tilde{S}_n(f_0, x)$ of degree $\le n$. Finally, we put

$$S_n^*(f,x) := \tilde{S}_n(f_0, x) + Q_r(f,x)\,. \tag{5.9}$$

This completes the construction of the S_n^*. They are linear bounded operators and satisfy

$$f(x) - S_n^*(f,x) = g_0(t) - S_n(g_0, t)\,, \qquad x = \cos t\,. \tag{5.10}$$

Assume first that $f \in C^r$. From §7 of Chapter 4, for some $\xi \in (-1,1)$, $|f(x) - Q_r(f,x)| \le C|f^{(r)}(\xi)| \le C\|f^{(r)}\|$. Thus, by the inequality (5.8) of Chapter 2

$$\|f_0^{(k)}\| \le C\|f_0^{(r)}\| = C\|f^{(r)}\|\,, \qquad k = 0, 2, \dots, r\,. \tag{5.11}$$

The polynomials Q_r have been introduced in order to eliminate the intermediate derivatives in the definition of M in Lemma 5.2; we can now replace M by $C\|f^{(r)}\|$.

For the functions h_k which correspond to f_0 in Lemma 5.2, we have by Lemma 5.1,

$$|h_k(t) - S_n(h_k, t)| \le C\|f^{(r)}\| n^{-(r+k)}(|\sin t|^{r-k} + n^{k-r})\,.$$

Therefore,

$$\begin{aligned} |f(x) - S_n^*(f,x)| &= |f_0(t) - S_n(g_0, t)| \\ &\le C\|f^{(r)}\| \sum_{k=0}^{r} \left(\left|\frac{\sin t}{n}\right|^{r-k} \frac{1}{n^{2k}} + \frac{1}{n^{2r}} \right)\,. \end{aligned} \tag{5.12}$$

Now

$$|\sin t| = \sqrt{1-x^2}\,, \quad \text{and} \quad \sum_{k=0}^{n} \left|\frac{\sin t}{n}\right|^{r-k} \frac{1}{n^{2k}} \le \left(\frac{|\sin t|}{n} + \frac{1}{n^2} \right)^r = \Delta_n(x)^r\,.$$

Therefore, the sum in (5.12) does not exceed $C\Delta_n(x)^r$. We obtain, for all $f \in C^r$,

$$|f(x) - S_n^*(f,x)| \le C\Delta_n(x)^r \|f^{(r)}\|\,.$$

We apply Theorem 5.3 of Chapter 7 to obtain the required relation (5.8). $\square$

In many cases, the quantity $\Delta_n(x)$ in the above estimates can be replaced by the smaller one

$$\delta_n(x) = \frac{\sqrt{1-x^2}}{n}\,, \qquad x \in A := [-1,1]\,, \quad n = 1, 2, \dots\,, \quad \delta_0(x) = 1\,.$$

We shall show later that Brudnyi's theorem cannot be improved in this way if $r \ge 3$. We prove here with DeVore [1976] that the improvement is possible if $r = 2$.

Theorem 5.4. *There exist operators S_n^{**} on $C(A)$ and an absolute constant C such that for all continuous f on A,*

$$|f(x) - S_n^{**}(f,x)| \le C\omega_2(f, \delta_n(x))\,, \qquad -1 \le x \le 1\,. \tag{5.13}$$

Proof. With the operators S_n^* of Theorem 5.3, for $r = 2$, we define S_n^{**} by $S_n^{**}(f) := S_n^*(f) - l(f)$, where $l(f,x)$ is the linear function which interpolates $f - S_n^*(f)$ at the points ± 1. Thus, $f(x) - S_n^{**}(f,x)$ vanishes at these points.

First, let $f \in C^2$, and let $M := \|f''\|$. From (5.8)

$$|(f - S_n^* f)(\pm 1)| \le CM\Delta_n(\pm 1)^2 = CMn^{-4}.$$

Hence $|l(f,x)|$ and also $|l'(f,x)|$ do not exceed $CMn^{-4} \le CM\Delta_n(x)^2$. This and (5.8) ensure that

$$|f(x) - S_n^{**}(f,x)| \le CM\Delta_n(x)^2, \qquad -1 \le x \le 1.$$

On the interval $A_n := [-\alpha_n, \alpha_n]$, $\alpha_n := \sqrt{1-n^{-2}}$ this is equivalent to

$$|f(x) - S_n^{**}(f,x)| \le CM\delta_n(x)^2. \tag{5.14}$$

This also holds when $x \in A \setminus A_n$, but is a little more difficult to prove. We go back to the function $g_0(t) = f_0(\cos t)$ of Lemma 5.2. Since $r = 2$, we have $g_0 = h_0 + h_1 + h_2$ with the h_k satisfying (5.5). For the derivative $g_0'' = h_0'' + h_1'' + h_2''$, we obtain, using (5.5) for the h_k'',

$$|S_n(g_0'', t) - g_0''(t)| \le CM(\sin^2 t + n^{-2}). \tag{5.15}$$

Now $S_n(g_0') - g_0' = S_n(g_0)' - g_0'$ is an odd function and therefore vanishes at $t = 0, \pi$. Integrating $S_n''(g_0) - g_0''$ on the shorter of the intervals $(0,t)$ and (t,π) we obtain

$$|S_n(g_0', t) - g_0'(t)| \le CM(|\sin t|^3 + n^{-2}|\sin t|).$$

If we differentiate $f(x) - S_n^{**}(f,x) = g_0(t) - S_n(g_0,t) - l(f,x)$ with respect to t, we obtain from this

$$\begin{aligned}|f'(x) - S_n^{**}(f,x)'| &\le CM\big(|\sin t|^{-1}(|\sin t|^3 + n^{-2}|\sin t|) + CMn^{-4}\big)\\ &\le CM(\sin^2 t + n^{-2}) \le CM(1 - x^2 + n^{-2}).\end{aligned} \tag{5.16}$$

In estimating $f(x) - S_n^{**}(f,x)$ for $x \in A \setminus A_n$, we can restrict ourselves by symmetry to the case $\alpha_n \le x \le 1$. Then, the last term in (5.16) is $\le CMn^{-2}$. Integrating this relation on $(x,1)$ yields

$$|f(x) - S_n^{**}(f,x)| \le CMn^{-2}(1-x) = CM\delta_n(x)^2.$$

We have proved (5.14) for $-1 \le x \le 1$. It remains only to apply Theorem 5.3 of Chapter 7. □

Theorem 5.4 does not extend to the moduli of smoothness of order $r > 2$ as the following example of X.M. Yu shows:

Example. For each constant $M > 0$, and each $n \ge 4$, there is a function $f_n \in C(A)$, $A := [-1,1]$, such that for each polynomial $P_n \in \mathcal{P}_n$

$$|f_n(x) - P_n(x)| \le M\omega_3(f_n, \delta_n(x)),$$

is violated for some $x \in [-1,1]$.

Proof. We can assume $M \ge 1$. For $n \ge 4$, let $f(x) := f_n(x) := (x - 1 + a)_+^3$, $x \in A$, where $a := M^{-1}n^{-2}$. Then, $\omega_3(f,h) \le 8\min(h^3, a^3)$, $h > 0$. We suppose that there is a polynomial P_n such that

$$|f(x) - P_n(x)| \le M\omega_3(f, \delta_n(x)) \le 8M\min(\delta_n(x)^3, a^3), \qquad x \in A, \tag{5.17}$$

and we derive a contradiction. Since $\|f\| \le a^3$, from (5.17), we conclude $\|P_n\| \le (8M+1)a^3 \le 9Ma^3$. Hence by Markov's inequality $\|P_n''\| \le 9Mn^4a^3$. Now, (5.17) implies that $f - P_n$ has a double zero at $x = 1$. Hence, $P_n(1) = f(1) = a^3$ and $P_n'(1) = f'(1) = 3a^2$. Therefore, from Taylor's remainder formula (see (5.6) of Chapter 2), we have for $x_0 := 1 - 10Ma$,

$$P_n(x_0) = a^3 - 30Ma^3 + \frac{1}{2}P_n''(\xi)(Ma)^2, \qquad \xi \in (1 - 10Ma, 1) \tag{5.18}$$

The last term in (5.18) does not exceed $5Mn^4a^3M^2a^2 = 5Ma^3$. Using (5.18) again, we find $|P_n(x_0)| \ge 24Ma^3$. But this contradicts (5.17) because $f_n(x_0) = 0$. □

§ 6. Inverse Theorems

Proofs of the theorems inverse to the results of §4-5 parallel those of §2 of Chapter 7. However, instead of the ordinary Bernstein inequality, Theorem 2.3 is required.

Theorem 6.1 (Dzyadyk, Timan, Brudnyi). *Let $\Phi \in \Omega_r$ for some $r = 1, 2, \ldots$. If a function $f \in C(A)$, $A := [-1,1]$ satisfies*

$$|f(x) - P_n(x)| \le \Phi(\Delta_n(x)), \qquad x \in A,\ n = 1, 2, \ldots \tag{6.1}$$

for some sequence of polynomials P_n of degree $\le n$, then

$$\omega_r(f,t) \le Ct^r \int_t^1 \Phi(s)s^{-r-1}\,ds, \qquad 0 < t < \tfrac{1}{4} \tag{6.2}$$

with C depending only on r and the constants in (2.8) *and* (2.9).

Proof. If $\Delta_t^r(f,x)$, $x \in A$, $t > 0$ is meaningful, then $J := [x, x+rt] \subset A$. Our first remark is that if a, b are the smallest and the largest value of $|u|$, $u \in J$, if $C_0 \ge 1$, and if for some $m = 1, 2, \ldots$,

$$|b - a| \le C_0\Delta_m(b), \tag{6.3}$$

then

$$\Delta_m(a) \le 4C_0\Delta_m(b). \tag{6.4}$$

We consider two cases. First let $\sqrt{1-b^2} \ge m^{-1}$. Then the statement (6.4) will follow if we show that $\sqrt{1-a^2} \le 4C_0\sqrt{1-b^2}$. The latter follows from the mean value theorem:

$$\sqrt{1-a^2}-\sqrt{1-b^2} \le \frac{|b-a|}{\sqrt{1-b^2}} \le C_0\frac{\Delta_m(b)}{\sqrt{1-b^2}} \le 2C_0\frac{\sqrt{1-b^2}}{m\sqrt{1-b^2}} = 2C_0m^{-1}.$$

On the other hand, if $\sqrt{1-b^2} \le m^{-1}$, we need only to show that $\sqrt{1-a^2} \le 3C_0m^{-1}$. This follows because $b^2-a^2 \le 2|b-a|$, hence $1-a^2 \le 1-b^2+2|b-a|$, so that

$$\sqrt{1-a^2}-\sqrt{1-b^2} \le \sqrt{2|b-a|} \le \sqrt{2C_0\Delta_m(b)} \le 2C_0m^{-1}.$$

The remaining part of the proof of (6.2) is standard, one uses a continuous analogue of (3.5) of Chapter 7.

Using (6.1), for $J \subset A$ we have, with some $\xi \in J$, from the remark following (2.10),

$$\begin{aligned} |\Delta_t^r(f,x)| &\le 2^r \max_{y\in J}|f(y)-P_{2^n}(y)| + |\Delta_t^r(P_{2^n},x)| \\ &\le 2^r\Phi(\Delta_{2^n}(a)) + t^r|P_{2^n}^{(r)}(\xi)|. \end{aligned} \tag{6.5}$$

Let k be the smallest integer with $2^k \ge r$. We use the abbreviations $a_j := \Delta_{2^j}(a)$, $b_j := \Delta_{2^j}(b)$, $Q_j = P_{2^j}^{(r)}$, for $j \ge k-1$; $Q_{k-1} = 0$ since $P_{2^{k-1}} \in \mathcal{P}_{r-1}$ (we also put $Q_{-1} = P_0$ in case $r=1$, $k=0$). Then

$$|P_{2^n}^{(r)}(\xi)| \le \sum_{j=k}^{n} |Q_j(\xi) - Q_{j-1}(\xi)|.$$

From (2.8)–(2.10) we see that $\Phi(b_{j-1}) \sim \Phi(b_j)$, $j \ge k-1$. From (6.1) and (2.8) we deduce $|P_{2^j}(x) - P_{2^{j-1}}(x)| \le C\Phi(\Delta_{2^j}(x))$. Moreover, from (2.9), $\Delta_{2^j}(\xi)^{-r}\Phi(\Delta_{2^j}(\xi)) \le Cb_j^{-r}\Phi(b_j)$, and so, by Theorem 2.3,

$$|Q_j(\xi) - Q_{j-1}(\xi)| \le C\Delta_{2^j}(\xi)^{-r}\Phi(\Delta_{2^j}(\xi)) \le Cb_j^{-r}\Phi(b_j).$$

Since $b_j \le b_{j-1} - b_j$,

$$b_j^{-r}\Phi(b_j) \le b_j^{-r-1}\Phi(b_j)(b_{j-1}-b_j) \le C\int_{b_j}^{b_{j-1}} \Phi(s)s^{-r-1}\,ds.$$

This yields

$$|\Delta_t^r(f,x)| \le C\left\{\Phi(a_n) + t^r\int_{b_n}^1 \Phi(s)s^{-r-1}\,ds\right\}. \tag{6.6}$$

For a given x with $J \subset A$ we select $n := n(x)$ so that $b_{n+1} < t \le b_n$; this is possible since $b_1 \ge \frac{1}{4} > t$. With this choice of n, $|b-a| \le rt \le rb_n$. We can use (6.4):

$$\Phi(a_n) \le C\Phi(b_n) \le Ct^r\int_{b_n}^1 \Phi(s)s^{-r-1}\,ds \le Ct^r\int_t^1 \Phi(s)s^{-r-1}\,ds.$$

Thus from (6.6) we have derived that $|\Delta_t^r(f,x)|$ and hence also that $\omega_r(f,t)$ has the required upper bound. □

In some cases, one can deduce the differentiability of f from estimates of the type (6.1).

Theorem 6.2 (Dzyadyk and Timan). *Let ω be a modulus of continuity with the property*

$$\int_0^1 \frac{\omega(s)}{s}\,ds < +\infty. \tag{6.7}$$

If for some function $f \in C(A)$ there exists a sequence of polynomials P_n for which

$$|f(x) - P_n(x)| \le \Delta_n(x)^r\omega(\Delta_n(x)), \qquad x \in A,\ n = 1,2,\ldots, \tag{6.8}$$

then $f \in C^r(A)$ and

$$\omega(f^{(r)},t) \le C_r\left\{t\int_t^1 \frac{\omega(s)}{s^2}\,ds + \int_0^t \frac{\omega(s)\,ds}{s}\right\}. \tag{6.9}$$

For the proof, see Timan [A-1963, p. 347].

Combining (6.1) with the results of §4 yields as the most important special case:

Theorem 6.3. *A function f belongs to the generalized Lipschitz class $\mathrm{Lip}^*\,\alpha$, $\alpha > 0$ if and only if $|f(x) - P_n(x)| \le M\Delta_n(x)^\alpha$, $x \in A$, $n = 1,2,\ldots$ for some M and some sequence P_n.*

§7. Approximation Spaces for Algebraic Polynomials

It is only recently that characterizations have been given for functions f with a specific order of the error of approximation by algebraic polynomials, such as $E_n(f)_p = \mathcal{O}(n^{-\alpha})$ on $[-1,1]$. The best approach to this question is by Ditzian and Totik (see their book [A-1987]). These characterizations require measurements of smoothness of f which depend upon the position of the point x in $A := [-1,1]$. For this purpose, one uses the weighted moduli of smoothnes, $\omega_r^\phi(f)$, with $\phi(x) := (1-x^2)^{1/2}$, of §6 of Chapter 6.

In this section, we first establish, for functions f in the Sobolev spaces $W_p^r(\phi)$ with weight ϕ of §6 of Chapter 6, the Jackson and the Bernstein type inequalities, namely:

(7.1) $$E_n(f)_p \le C_r n^{-r} \|\phi^r f^{(r)}\|_p\,, \qquad n \ge r$$

(7.2) $$\|\phi^r P_n^{(r)}\|_p \le C_r n^r \|P_n\|_p\,, \quad P \in \mathcal{P}_r\,, \qquad n = 1, 2, \dots$$

(compare §5 of Chapter 7).

From this we derive an estimate of the error $E_n(f)_p$ in terms of $\omega_r^\phi(f)$ in Theorem 7.3, and finally, in Theorem 7.7, the desired characterization of the approximation classes.

We begin by proving (7.1); we follow DeVore and Scott [1984]. A simple observation is that it will be sufficient to derive (7.1) for functions f with an absolutely continuous derivative $f^{(r-1)}$ on A. Indeed, if $f \in W_p^r(\phi)$, then, under this assumption, the functions $f_\rho(x) := f(\rho x)$, $\rho < 1$, converge to f in L_p for $\rho \to 1$ and satisfy $\|\phi^r f_\rho^{(r)}\|_p \le \|\phi^r f^{(r)}\|_p$. Hence, from (7.1) for the f_ρ we obtain (7.1) for f.

We represent the function f by means of

(7.3) $$f(x) = T(x) + \frac{1}{r!} \int_A f^{(r)}(t)(x-t)_+^{r-1}\,dt\,,$$

where T is the Taylor polynomial of degree $r-1$ of f at -1. We fix n and construct polynomials $Q_t(x)$, $t \in A$, of degree at most $4rn$, which approximate $(x-t)_+^{r-1}$, then our approximation to f, also a polynomial of degree $\le 4rn$, will be given by

(7.4) $$Q(x) = T(x) + \frac{1}{r!} \int_A f^{(r)}(t) Q_t(x)\,dt\,.$$

Let C_n be the Chebyshev polynomial of degree n and let $t_k = \cos\theta_k$, $\theta_k := (2k-1)\pi/2n$, $k = 1, \dots, n$, be its zeros. For convenience, let $t_0 := 1$, $t_{n+1} := -1$. We shall use the kernels

(7.5) $$\Lambda_k(x) := c_k \left(\frac{C_n(x)}{x - t_k} \right)^{4r}, \qquad k = 1, 2, \dots, n$$

with c_k chosen so that $\int_A \Lambda_k(x)\,dx = 1$. (The operator (7.4) is related to the operator (3.6) of Chapter 9 of Bojanic and DeVore [1969].) With this, we define

(7.6) $$Q_t(x) := \begin{cases} (x-t)^{r-1}\,, & \text{if } t \le t_n\,, \\ 0\,, & \text{if } t > t_1 \\ (x-t)^{r-1} \int_{-1}^x \Lambda_k(u)\,du\,, & \text{if } t_{k+1} < t \le t_k\,, \quad k = 1, \dots, n-1\,. \end{cases}$$

We shall need some simple estimates of the constants c_k and of the functions $\delta_n(x) := \phi(x)/n = \sqrt{1-x^2}/n$. First of all, for $t \in I_k := [t_{k+1}, t_k]$, for large n,

(7.7) $$\tfrac{1}{3}\delta_n(t_k) \le \delta_n(t) \le 3\delta_n(t_k)\,, \qquad k = 1, \dots, n-1\,.$$

To see this, let $t = \cos\theta$, $-1 \le t \le 1$, then $n\delta_n(t) = |\sin\theta|$. By symmetry, we can assume that $t_k > 0$. Let also $t_{k+1} > 0$. Because $\sin\theta/\theta$ decreases for $0 \le \theta \le \frac{\pi}{2}$, we have

$$n\delta_n(t_{k+1}) = \sin\theta_{k+1} \le \frac{\theta_{k+1}}{\theta_k} \sin\theta_k \le \frac{\theta_2}{\theta_1} \sin\theta_k = 3n\delta_n(t_k)\,.$$

Since $\delta_n(t)$ decreases on [0,1], we get (7.7). If $t_{k+1} < 0$, we compare in the same way $\delta_n(t_k)$ with $\delta_n(0) = \max \delta_n(t)$.

We also have

(7.8) $$\begin{cases} t_k - t_{k+1} \le 3\pi\delta_n(t_k)\,, & k = 1, \dots, n-1 \\ 1 - t_1 = t_n + 1 \le \pi\delta_n(t_1)\,. \end{cases}$$

For example, the first of these inequalities, because of (7.7), follows from $\cos\theta_k - \cos\theta_{k+1} = (\pi/n)|\sin\xi| \le 3\pi\delta_n(t_k)$, for some $\xi \in I_k$. We have also the (uniform in k) weak equivalences $t_k - t_{k+1} \sim \delta_n(t_k)$ and then, because of (7.7), the relation $t_k - t_{k+1} \sim t_{k-1} - t_k$. As a consequence, let an interval I_j be contained between $x, t \in A$. Then, if $t \in I_k$,

(7.9) $$|x - t| \le C|x - t_k|\,,$$

with an absolute constant C. This is obvious if $x < t$; if $x > t$, it follows from

$$\begin{aligned} x - t &\le t_k - t_{k+1} + t_{k-1} - t_k + x - t_{k-1} \\ &\le C(t_{k-1} - t_k) + x - t_{k-1} \le C(x - t_k)\,. \end{aligned}$$

Lemma 7.1. *With a constant $C > 0$ depending only on r, we have*

(7.10) $$c_k \le C\delta_n(t_k)^{4r-1}\,, \qquad k = 1, \dots, n\,.$$

Proof. For a given k, we consider the interval

$$I := \left[\frac{k\pi}{n} - \frac{\pi}{6n}, \frac{k\pi}{n} + \frac{\pi}{6n} \right] \subset I_k\,.$$

For $\theta \in I$, $|\cos n\theta| \ge \frac{1}{2}$ and with some $t \in I_k$, $|\sin\theta| = n\delta_n(t) \ge \frac{1}{3} n\delta_n(t_k)$. By means of the substitution $t = \cos\theta$ we obtain

(7.11) $$c_k^{-1} \int_{-1}^{+1} \Lambda_k(t)\,dt \ge \int_I \frac{|\cos n\theta|^{4r}\,|\sin\theta|\,d\theta}{|\cos\theta - \cos\theta_k|^{4r}}\,.$$

Here, by the mean value theorem and (7.7), with $\xi \in I$,

$$|\cos\theta - \cos\theta_k| = |\theta - \theta_k|\,|\sin\xi| \le \frac{\pi}{n}|\sin\xi| \le C\delta_n(t_k)\,.$$

Therefore, the integral (7.11) is

$$\ge C \frac{n|I|}{\delta_n(t_k)^{4r-1}} \ge C\delta_n(t_k)^{-4r+1}\,,$$

which yields (7.10). □

We can now estimate the error of approximation $E(x,t) := E_r(x,t) := |(x-t)_+^{r-1} - Q_t(x)|$. Since $\int_A \Lambda_k(u)\,du = 1$, we have

$$E(x,t) \le C|x-t|^{r-1}, \quad \text{for } x,t \in A. \tag{7.12}$$

In some cases this can be improved. We have, if $t \in I_k$,

$$E_0(x,t) = |(x-t)_+^0 - \int_{-1}^{x} \Lambda_k(u)\,du| \le \begin{cases} \int_{-1}^{x} \Lambda_k(u)\,du, & x < t, \\ \int_{x}^{1} \Lambda_k(u)\,du, & x \ge t. \end{cases}$$

In the case that $t \in I_k$, $x \notin I_k$, each of the integrals on the right side does not exceed

$$c_k \int_{|x-t_k|}^{\infty} u^{-4r}\,du \le c_k \frac{|x-t_k|^{-4r+1}}{4r-1}.$$

From this and (7.10), we obtain for $k = 1,2,\dots,n-1$:

$$E_r(x,t) = |x-t|^{r-1} E_0(x,t) \le CM(x,t), \quad t \in I_k, \quad x \notin I_k, \tag{7.13}$$

where

$$M(x,t) := |x-t|^{r-1}|x-t_k|^{-4r+1}\delta_n(t_k)^{4r-1}, \quad t \in I_k. \tag{7.14}$$

We also note that

$$E_r(x,t) = 0 \ \text{ if } x \le t, \ t \ge t_1 \ \text{ or if } x \ge t, \ t \le t_n. \tag{7.15}$$

The following is our key technical lemma.

Lemma 7.2. *With a constant $C > 0$ that depends only on $r = 1,2,\dots$ we have*

$$\begin{cases} \text{(i)} & \int_A E_r(x,t)\delta_n(t)^{-r}\,dx \le C, \quad t \in A \\ \text{(ii)} & \int_A E_r(x,t)\delta_n(t)^{-r}\,dt \le C, \quad x \in A. \end{cases} \tag{7.16}$$

Proof. Because of symmetry, we need to prove (i) or (ii) only for $t \in [0,1]$ or for $x \in [0,1]$, respectively.

(i). If $t \in [t_1, 1]$, then by (7.15) and (7.12), and using $1 - t \le Cn^{-2}$, we see that the integral (i) does not exceed $C\delta_n(t)^{-r}(1-t)^r \le Cn^r(1-t)^{r/2} \le C$.

It remains to consider the case $t \in I_k$, for some fixed $k = 1,\dots,n-1$, $t \ge 0$. We split the integration range into $|t-x| \le \delta$ and $|t-x| > \delta$, taking $\delta = \lambda\delta_n(t_k)$ and the constant λ so large that $[t_{k+2}, t_{k-1}]$ is contained in the range of $|t-x| \le \delta$. This is possible, with an absolute constant (independent of k), because $\delta_n(t_k)$, $|I_k| = t_k - t_{k+1}$ and the lengths of both intervals adjoining I_k are uniformly weakly equivalent. Then $|t-x| \ge \delta$ will imply that $x \notin I_k$,

and that there is an interval I_j between t and x. For the part of the integral over $|x-t| \le \delta$ we apply (7.12) and get the upper bound $C\delta_n(t_k)^{-r}\delta^r \le C$. In the remaining integral, we replace $E(x,t)$ with $CM(x,t)$ by (7.14), $\delta_n(t)$ with $\delta_n(t_k)$ by (7.7), $|x-t_k|$ with $|x-t|$ by (7.9). We obtain for the integral the upper bound $C\delta_n(t_k)^{3r-1}\int_{|x-t|\ge\delta}|x-t|^{-3r}\,dx$, which does not exceed a constant since

$$\int_{u\ge\delta}^{\infty} u^{-3r}\,du = \frac{\delta^{-3r+1}}{3r-1}.$$

(ii). Similar computations are possible for (ii); we can assume that $x \ge 0$. First we estimate the integral over $[t_1, 1]$ (and by symmetry, over $[-1, t_n]$). By (7.15), it is zero unless $x > t_1$, in which case it does not exceed, by (7.12) and the definition of δ_n,

$$C\int_{t_1}^{1} |x-t|^{r-1}\delta_n(t)^{-r}\,dt \le Cn^r \int_{t_1}^{1} \frac{(1-t)^{r-1}}{(1-t)^{r/2}}\,dt \le Cn^r(1-t_1)^{r/2} \le C.$$

It remains to estimate $K := \int_{t_n}^{t_1} E\delta_n^{-r}\,dt$ for $x \in I_k$, with fixed $k = 0,\dots,n$. We define $\delta := \lambda\delta_n(t_k)$ as in (i), and split K into parts with ranges of integration $|t-x| \le \delta$ and $|t-x| \ge \delta$. As before, the first part is bounded. For the second part, we have to estimate $\int_{|t-x|\ge\delta} M(x,t)\,dt$, with $M(x,t)$ from (7.14). Using again (7.7) and (7.9), we get that

$$M(x,t) \le C|t-x|^{-3r}\delta_n(t)^{3r-1}, \quad |t-x| \ge \delta, \quad t_1 \le t \le t_n.$$

Therefore this part of K does not exceed

$$C\left(\int_{x+\delta}^{1} + \int_{-1}^{x-\delta}\right)|t-x|^{-3r}\delta_n(t)^{3r-1}\,dt. \tag{7.17}$$

For $x \ge 0$, $\delta_n(t)$ decreases for $t \ge x$, and the first integral is $\le C\delta^{3r-1}\int_{x+\delta}^{1}(t-x)^{-3r}\,dt \le C$.

The estimate of the second integral, which is equal to $\int_{\delta}^{x+1} u^{-3r}\delta_n(x-u)^{3r-1}\,du$, is slightly more difficult. Here we have

$$\begin{aligned} n\delta_n(x-u) &= \sqrt{1-(x-u)^2} \le 2\sqrt{1-x+u} \\ &\le 2(\sqrt{1-x} + \sqrt{u}) \le 2n\delta_n(x) + 2\sqrt{u}. \end{aligned}$$

Therefore the integral is not more than

$$C\delta_n(x)^{3r-1}\delta^{-3r+1} + Cn^{-3r+1}\delta^{-\frac{3r}{2}+\frac{1}{2}}. \tag{7.18}$$

But $\delta_n(x) \le C\delta$ and $\delta = \lambda\delta_n(t_k) \ge Cn^{-2}$ so that (7.18) does not exceed a constant. □

Theorem 7.3. *For each $r = 1,2,\dots$, the inequality (7.1) holds for $n \ge r$. Moreover, for $f \in L_p(I)$, we have*

$$E_n(f)_p \le C_r\omega_r^{\phi}(f, 1/n)_p, \qquad n \ge 4r \tag{7.19}$$

where $\omega_r^{\phi}(f,t)$ is the weighted modulus of smoothness of f.

Proof. Using the definitions of $Q(x)$, of $E(x,t)$ and the relation between ϕ and δ_n, we get

$$\begin{aligned} |f(x) - Q(x)| &\le \frac{1}{r!}\int_A |(x-t)_+^{r-1} - Q_t(x)|\,|f^{(r)}(t)|\,dt \\ &= n^{-r}\int_A K(x,t)\phi(t)^r |f^{(r)}(t)|\,dt \end{aligned}$$

with the kernel $K(x,t) := \frac{1}{r!}E(x,t)\delta_n(t)^{-r}$. The inequalities (7.16) are exactly the conditions of Theorem 4.5 of Chapter 2 for K. Thus, the operator mapping a function $g(x)$ onto $\int_A K(x,t)g(t)\,dt$ is bounded on any rearrangement invariant space X on A. This yields

$$\|f - Q\|_X \le Cn^{-r}\|\phi^r f^{(r)}\|_X\,.$$

In particular, for $X = L_p$ we obtain $E_{4nr}(f)_p \le Cn^{-r}\|\phi^r f^{(r)}\|_p$, hence also (7.1).

Next, we use the abstract Theorem 5.1, (i) of Chapter 7. In our situation, (7.1) is the Jackson inequality (5.4) of Chapter 7 if $X := L_p$, $Y := W_p^r(\phi)$ and $\Phi_n := \mathcal{P}_n$. The K-functional $K(f,t^r; L_p, W_p^r(\phi))$ is equivalent to $\omega_r^\phi(f,t)$ by Theorem 6.2 of Chapter 6. In this way, we get (7.19). □

We shall need some new inequalities for trigonometric polynomial T_n on $\mathbb{T}$.

Lemma 7.4. *Let $0 < p \le \infty$. For each trigonometric polynomial T_n of degree $\le n$,*

$$\|T_n\|_p \le Cn\|\sin t\; T_n(t)\|_p\,, \qquad n = 1,2,\dots, \tag{7.20}$$

where C is an absolute constant.

Proof. Let $B = \{t \in \mathbb{T} : |\sin t| \le \sin\frac{1}{8n}\}$; the set B consists of two intervals of length $\frac{1}{4n}$. On $B_1 := \mathbb{T}\setminus B$, $n|\sin t| \ge \frac{1}{4\pi}$, therefore

$$\|T_n\|_p(B_1) \le 4\pi n\|\sin t\; T_n(t)\|_p(\mathbb{T})\,. \tag{7.21}$$

To estimate $\|T_n\|_p(B)$, we let $M := \|T_n\|_\infty(\mathbb{T})$, then

$$\|T_n\|_p(B) \le M|B|^{1/p}\,. \tag{7.22}$$

Let $t_0 \in \mathbb{T}$ be a point where $|T_n(t_0)| = M$, and let I be the interval $I := [t_0 - \frac{1}{2n}, t_0 + \frac{1}{2n}]$. By Bernstein's inequality, $|T_n(t) - T_n(t_0)| \le \frac{1}{2n}\|T_n'\|_\infty \le \frac12 M$, $t \in I$, hence $|T_n(t)| \ge M/2$ on this interval. Now the set $B_2 := I \cap B_1$ has measure $|B_2| \ge \frac1n - \frac{1}{2n} = |B|$, and therefore, by (7.22),

$$\begin{aligned} n\|\sin t\; T_n(t)\|_p(\mathbb{T}) &\ge n\|\sin t\; T_n(t)\|_p(B_2) \ge \frac{1}{4\pi}\,\frac{M}{2}|B_2|^{1/p} \\ &\ge \frac{1}{8\pi}M|B|^{1/p} \ge \frac{1}{8\pi}\|T_n\|_p(B)\,. \end{aligned}$$

It follows that $\|T_n\|_p(B) \le 8\pi n\|\sin t\; T_n(t)\|_p$, and together with (7.21), we obtain (7.20) with $C = 12\pi$. □

Theorem 7.5 (Ditzian and Totik [A-1988]). *Let $0 < p \le \infty$, $r = 1,2,\dots$. Then for any trigonometric polynomial T_n of degree $\le n$,*

$$\|\sin^r t\; T_n'(t)\|_p \le Cn\|\sin^r t\; T_n(t)\|_p\,, \qquad n = 0,1,\dots, \tag{7.23}$$

with the constant C depending only on r and on p, if p is close to zero.

Proof. We can assume $n \ne 0$. To the relation

$$\sin^r t\; T_n'(t) = \big(\sin^r t\; T_n(t)\big)' - r\sin^{r-1} t\cos t\; T_n(t)$$

we apply the $L_p(\mathbb{T})$ (quasi-) norm and see that the left side of (7.23) does not exceed

$$C_p\big\{\|(\sin^r t\; T_n(t)'\|_p + r\|\sin^{r-1} t\; T_n(t)\|_p\big\}\,.$$

We use Bernstein's inequality for the first term, (7.20) for the second, completing the proof. □

Theorem 7.6. *Inequality (7.2) holds for $1 \le p \le \infty$ and $n = 0,1,\dots$. Moreover, for $m = 1,2,\dots$ and any polynomial P_n of degree $\le n$, we have*

$$\|\phi^m P_n'\|_p \le C_m n\|\phi^{m-1}P_n\|_p\,. \tag{7.24}$$

Proof. Let P_n be an algebraic polynomial of degree $\le n$ and let $T_n(t) := P_n(x)$, $x = \cos t$, then $T_n'(t) = P_n'(x)\sin t$. From (7.23) we get

$$\|\phi^m P_n'\|_p = \|\sin^m t\, T_n'(t)\|_p \le Cn\|\sin^m t\, T_n(t)\|_p = Cn\|\phi^{m-1}P_n\|_p\,.$$

A repeated application of the last inequality yields (7.2). □

We can now apply Theorem 5.1, (ii) of Chapter 7 and obtain the inverse estimate

$$\omega_r^\phi(f,1/n)_p \le Cn^{-r}\sum_{k=r}^n k^{r-1}E_k(f)_p\,, \qquad n > r\,. \tag{7.25}$$

The companion inequalities (7.19) and (7.25) serve to characterize functions with a prescribed order of approximation. We mention the following special cases:

Theorem 7.7. *Let $1 \le p \le \infty$, $0 < q \le \infty$ and $0 < \alpha < r$. For $f \in L_p(I)$, we have*

(i) $E_n(f)_p = \mathcal{O}(n^{-\alpha})$, $n = 1,2,\dots$,
if and only if $\omega_r^\phi(f,t)_p = \mathcal{O}(t^\alpha)$, $t > 0$.

(ii) $\sum_{n=1}^\infty \big[n^\alpha E_n(f)_p\big]^q n^{-1} < \infty$ *if and only if* $\int_0^\infty \big[t^{-\alpha}\omega_r^\phi(f,t)_p\big]^q\,dt/t < \infty$.

As examples we mention for $\alpha, \beta \ge 0$ the functions

$$f_1(x) = (1-x)^\alpha\,, \quad \alpha \ne 0, 1, \dots\,; \quad f_2(x) = (1-x)^\alpha \log^\beta (1-x)\,, \quad x \in A\,,$$

with errors of polynomial approximation

$$E_n(f_1)_p = \mathcal{O}(n^{-2\alpha - 2/p})\,,$$

$$E_n(f_2) = \begin{cases} \mathcal{O}(n^{-2\alpha-2/p}(\log n)^\beta)\,, & \alpha \ne 0, 1 \\ \mathcal{O}(n^{-2\alpha-2/p}(\log n)^{\beta-1})\,, & \alpha = 0,\ \alpha = 1,\ \beta > 0\,. \end{cases} \tag{7.26}$$

This follows by computing $\omega_r^\phi(f_1)$, $\omega_r^\phi(f_2)$ (see Problem 8.2 of Chapter 6).

§ 8. Problems and Notes

8.1. The quantity $\delta_n(x)$ of (1.1) is strongly equivalent to the distance from x to the Bernstein ellipse E_ρ, $\rho = 1 + \frac{1}{n}$, at least if $|x| \le 1 - (2n^2 + 2n + 1)^{-1}$, see Dzyadyk [A-1977, p. 258].

8.2. The L_p analogue of Timan's result (1.3) is not valid: Motornyi [1971], see also DeVore [1976]. That is, the class of functions $\mathrm{Lip}(\alpha, L_p)$ is not characterized by the condition $\|\Delta_n^{-\alpha}(f - P_n)\|_p = \mathcal{O}(1)$, $n \to \infty$. There are more complicated characterizations of $\mathrm{Lip}(\alpha, L_p)$ by algebraic polynomial approximation (Motornyi [1971], Oswald [1978]); see the paper of Oswald for a statement of these results.

8.3. In several papers, and in Chapter IX of his book [A-1977], Dzyadyk has developed a theory of approximation of functions on compact subsets of $\mathbb{C}$ which contains many results of this chapter. Let $A \subset \mathbb{C}$ be a compact set with connected complement, let $w = G(z)$ be a single valued analytic function, properly normed, which maps this complement onto the exterior of the circle $|w| = 1$. He finds polynomials P_n with small difference $|f(z) - P_n(z)|$, $z \in \partial A$. This difference is estimated in terms of the smoothness of f and of its derivatives on the boundary ∂A and of the distance $\Delta_n(z)$ of $z \in \partial A$ to the level curve $|G(z)| = 1 + \frac{1}{n}$. Much depends also on the smoothness properties of the curve ∂A. This explains the geometric meaning of the quantities $\Delta_n(x)$, $\delta_n(x)$ in §§4-5.

8.4. Also the inequality (2.15) of Lebed and Brudnyi has a complex analogue in the book Ditzian [A-1977]. This provides an alternate, albeit not simple approach to (2.15).

Chapter 9. Approximation by Operators

§ 1. Introduction

The operator of best approximation is continuous, but not linear (see §1 of Chapter 3), hence not easy to handle. But its main defect is the absence of simple formulæ for its representation. One is led therefore to simple expressions – for instance linear integral operators – that provide good, but not necessarily best approximation. We have studied the norms of integral operators in §4 of Chapter 2, also many special operators, for example the Jackson operators in §2 of Chapter 7. Operators will be the main subject of the present chapter.

A continuous linear operator U, mapping a Banach space X into itself, is of finite rank n if its range $X_n = \{Uf : f \in X\}$ is a linear space of dimension n. If the vectors $g_1, \dots, g_n$ span X_n, we can write

$$Uf = \sum_{i=1}^n c_i(f) g_i\,. \tag{1.1}$$

Applying to both sides of (1.1) a linear functional which is orthogonal to all but one g_i, one proves that the coefficients $c_i(f)$ are continuous linear functionals for $f \in X$.

A Banach space X is said to have the approximation property AP, if the identity operator I on X can be approximated, on each compact set $A \subset X$, by operators U_n of finite rank:

$$\|f - U_n f\| \to 0 \quad \text{for} \quad n \to \infty \quad \text{and all} \quad f \in A\,. \tag{1.2}$$

In this book, we prove that standard spaces of functions have AP. However, in 1973 Enflo constructed a separable Banach space without AP. See Davie [1975]. Approximation theory is hardly possible in such spaces.

The *norm* of a linear operator U which maps a Banach space X into another space Y is given by

$$\|U\| := \sup_{\|f\|_X \le 1} \|U(f)\|_Y\,; \tag{1.3}$$

U is continuous if and only if $\|U\| < +\infty$. (Then U is called *bounded*.) The importance of the norm lies in the *Banach-Steinhaus theorem*, which gives necessary and sufficient conditions for the convergence

$$(1.4)\qquad U_n(f) \to U(f)\,, \qquad f \in X\,,$$

of a sequence of linear bounded operators U_n. The conditions are:

(a) Convergence $U_n(g) \to U(g)$ for a set of $g \in X$ whose linear closure is X;
(b) Boundedness of the norms, $\|U_n\| \le M$, $n = 1, 2, \dots$.

In some cases, for positive operators U_n, (see Theorem 3.1 in Chapter 1), one has to test (a) only for a finite set of g. We shall study positive operators in §4.

An operator U of X into itself is a *projection* onto a linear supspace Y, if $Uf \in Y$ for all $f \in X$, and $Uf = f$ for $f \in Y$. If U_n are projections of X onto finite dimensional subspaces X_n, $n = 1, 2, \dots$, $X_1 \subset X_2 \subset \cdots$, and if $\overline{\cup X_n} = X$, then in the Banach-Steinhaus theorem one has to check only (b). Norms of projections appear in Lebesgue's lemma of §4, Chapter 2.

§ 2. Computation of Some Norms

We begin with properties of the operators $s_n(f, x)$ of partial sums of Fourier series, and of Fejèr means $\sigma_n(f, x)$ given by $\sigma_n(f) = \frac{1}{n+1}[s_0(f) + \cdots + s_n(f)]$ for $f \in X$. The space X *in this section will be a rearrangement-invariant Banach function space on* $\mathbb{T}$ (see Chapter 2). We recall some essential properties of these spaces. One has the continuous embeddings $C(\mathbb{T}) \subset X \subset L_1(\mathbb{T})$, with $C(\mathbb{T})$ dense in X. Other facts are that $\|f(x + a)\| = \|f\|$, $f \in X$; moreover, that $|g| \le |f|$, $f \in X$ implies $g \in X$ and $\|g\| \le \|f\|$.

Both $s_n(f, x)$ and $\sigma_n(f, x)$ are given by convolutions of the type

$$(2.1)\qquad \int_{\mathbb{T}} K_n(x - t) f(t)\, dt$$

where K_n is the continuous kernel $\frac{1}{\pi} D_n$ or $\frac{1}{\pi} F_n$, respectively (see §1 of Chapter 1). By Corollary 4.6 of Chapter 2, s_n and σ_n are bounded linear operators mapping each space X into itself. For the operator σ_n we have in addition $K_n \ge 0$, $\int_{\mathbb{T}} K_n\, dx = 1$. By the same Corollary 4.6, for all X,

$$(2.2)\qquad \|\sigma_n\|_X = 1\,.$$

We want to know whether s_n, σ_n have the *norm convergence properties* $s_n(f) \to f$, $\sigma_n(f) \to f$ on X. This is true for each trigonometric polynomial T_m. Indeed, $s_n(T_m) = T_m$, $n \ge m$; this also implies $\sigma_n(T_m) \to T_m$, $n \to \infty$. Since the trigonometric polynomials are dense in X, (2.2) and the Banach-Steinhaus theorem imply that $\|f - \sigma_n(f)\|_X \to 0$, $f \in X$. In particular:

1. *The Enflo phenomenon cannot happen in any of the rearrangement-invariant Banach function spaces on* $\mathbb{T}$.

In a similar way we have:

2. *A space X has the norm convergence property of the s_n if and only if the norms of the operators s_n are uniformly bounded.*

For $f \in X = L_2(\mathbb{T})$, from Parseval's identity, $\|s_n(f)\|_2 = \|f\|_2$, so that $\|s_n\|_{L_2(\mathbb{T})} = 1$. For the spaces $L_1(\mathbb{T})$, $C(\mathbb{T})$, one can evaluate the norms $\|s_n\|$ by direct computation.

Theorem 2.1 (Fejèr). *One has*

$$(2.3)\qquad \|s_n\|_p = \frac{4}{\pi^2} \log n + \mathcal{O}(1)\,, \qquad p = 1\,,\ p = \infty\,.$$

Proof. Since s_n is given by the integral (2.1) with $K_n = \frac{1}{\pi} D_n$, from Theorem 4.2 of Chapter 2 we see that each of the norms (2.3) is equal to

$$\Lambda_n := \frac{1}{\pi} \int_{\mathbb{T}} |D_n(t)|\, dt = \frac{2}{\pi} \int_0^\pi |D_n(t)|\, dt\,.$$

Since

$$2D_n(t) = \cot\frac{t}{2} \sin nt + \cos nt = \frac{2}{t} \sin nt + \left(\cot\frac{t}{2} - \frac{2}{t}\right) \sin nt + \cos nt$$

and $\cot u - u^{-1}$ is bounded on $(-\pi/2, \pi/2)$, we have $D_n(t) = \frac{\sin nt}{t} + \mathcal{O}(1)$. Therefore

$$\Lambda_n = \frac{2}{\pi} \int_0^\pi \frac{|\sin nt|}{t}\, dt + \mathcal{O}(1)\,.$$

The integral of $t^{-1}|\sin nt|$ over $(0, \frac{\pi}{n})$ is bounded, since $|\sin nt| \le nt$. Thus,

$$\begin{aligned}\Lambda_n &= \frac{2}{\pi} \sum_{k=1}^{n-1} \int_{k\pi/n}^{(k+1)\pi/n} \frac{|\sin nt|}{t}\, dt + \mathcal{O}(1) \\ &= \frac{2}{\pi} \int_0^{\pi/n} \sin nt \sum_{k=1}^{n-1} \frac{1}{t + n^{-1} k\pi}\, dt + \mathcal{O}(1)\,.\end{aligned}$$

Let $S(t)$ denote the last sum. For $0 \le t \le \pi/n$, $S(t)$ lies between

$$S(0) = n\pi^{-1} \left(1 + \frac{1}{2} + \cdots + \frac{1}{n-1}\right) \quad \text{and} \quad S(\pi/n) = S(0) + \mathcal{O}(n)\,.$$

We see that (2.3) follows from

$$1 + \frac{1}{2} + \cdots + \frac{1}{n-1} = \log n + \mathcal{O}(1)\,,$$

and

$$\int_0^{\pi/n} \sin nt\, dt = 2/n\,.$$

□

For functions $f \in L_1(\mathbb{T})$ it is convenient to have their Fourier series in the complex form, $f = \sum_{n=-\infty}^{+\infty} c_n(f) e_n$, with $c_n(f) := (1/2\pi) \int_{\mathbb{T}} f(x) e_{-n}(x)\, dx$, $e_n(x) := \exp(inx)$, $n \in \mathbb{Z}$. Then $s_n(f) = \sum_{k=-n}^{n} c_k(f) e_k$; the series

$-i\sum_{n=-\infty}^{\infty}(\operatorname{sign} n)c_n(f)e_n$ is the *conjugate Fourier series* of f (with sign 0 := 0). It can happen that it is also the Fourier series of an L_1 function $\tilde{f}$; then $\tilde{f}$ is the *conjugate function* of f. (In §3 of Chapter 11 we shall have another, more general definition of $\tilde{f}$.)

The space X on $\mathbb{T}$ has the *conjugation property* if for each $f \in X$, $\tilde{f}$ is defined and belongs to X. This gives us the operator $\tilde{U}(f) := \tilde{f}$ on X.

3. *If X has the conjugation property, then the operator $\tilde{U}$ is linear and bounded.*

Indeed, linearity of $\tilde{U}$ is obvious. The proof is completed by the use of the closed graph theorem. We have to show that $g_n \to f$ and $\tilde{g}_n = \tilde{U}(g_n) \to F$ in X imply $F = \tilde{f}$. For this, we compute the Fourier coefficients of F:

$$\begin{aligned} c_k(F) &= \lim_{n\to\infty} c_k(\tilde{g}_n) = -i(\operatorname{sign} k)\lim_{n\to\infty} c_k(g_n) \\ &= -i(\operatorname{sign} k)c_k(f) = c_k(\tilde{f})\,, \quad k \in \mathbb{Z}\,. \end{aligned}$$

Related to $\tilde{U}$ and $\tilde{f}$ are the operator U^* and the function f^*,

$$U^*f := f^* := \frac{1}{2}c_0(f) + \frac{1}{2}(f + i\tilde{f}) = \sum_0^\infty c_n(f)e_n\,.$$

Because of this formula and $\tilde{f} = -i(2f^* - f - c_0(f))$, we have

4. *The operator $\tilde{U}$ is well defined on X and is linear and bounded if and only if U^* has these properties.*

We can now prove with Katznelson [B-1968, p. 49]:

Theorem 2.2. *The rearrangement-invariant Banach function space X has bounded norms $\|s_n\|$ of the Fourier series if and only if it has the conjugation property.*

Proof. We note that for the partial sums $s_n^*(f,x) := \sum_{k=0}^n c_k(f)e_k(x)$ one has

$$\|s_{2n}^*\|_X = \|s_n\|_X\,. \tag{2.4}$$

In the space X, $|f| = |g|$, $f, g \in X$ imply that $\|f\| = \|g\|$. Therefore, (2.4) follows from

$$s_{2n}^*(f) = e_n s_n(e_{-n}f)\,.$$

(a) Let $\|s_n\| \le M$, $n = 0, 1, \ldots$, then the operators $s_n^*(f)$ are uniformly bounded; they converge on the set of all trigonometric polynomials in X. By one of the forms of the Banach-Steinhaus theorem, they converge on X to a linear bounded operator U. For each f, and k, the k-th Fourier coefficient of $U(f)$ is the limit of those of $s_n^*(f)$, that is, it is $= c_k(f)$, $k \ge 0$ or $= 0$ for $k < 0$. Thus, $U^* = U$ is defined and bounded on X.

(b) Conversely, let U^* be defined, linear and bounded on X. From

$$s_n^*(f) = U^*(f) - e_{2n+1}U^*(e_{-2n-1}f)$$

it follows that $\|s_n^*\|_X \le 2\|U^*\|$, so that the $\|s_n^*\|$, and then also the $\|s_n\|$ are bounded. □

We shall now use, without proof, the theorem of M. Riesz (see Zygmund [B-1959, vol. 1, p. 253]), according to which $L_p(\mathbb{T})$, $1 < p < \infty$, has the conjugation property:

$$\|\tilde{f}\|_p \le C_p\|f\|_p\,, \qquad 1 < p < \infty\,. \tag{2.5}$$

Corollary 2.3. *The norms of the operators $s_n(f)$ are bounded in each space $L_p(\mathbb{T})$, $1 < p < \infty$.*

From Lebesgue's Lemma 4.1 of Chapter 2 we now derive

$$\|f - s_n(f)\|_p = \begin{cases} C\log n\ E_n(f)_p, & p = 1, \infty\,,\ \ f \in C\,,\ \text{ if } p = \infty\,, \\ C_p\ E_n(f)_p, & 1 < p < \infty\,. \end{cases}$$

We see that the partial sums $s_n(f)$ approximate f almost as well as its polynomial of best approximation. This is true even for $f \in C$, if the factor $\log n$ is not essential for the problem considered.

We shall briefly discuss the norm of the Lagrange interpolation formula for algebraic polynomials

$$L_{n-1}(f,x) = \sum_{k=1}^n l_{n,k}(x)f(x_k) \tag{2.6}$$

with given interpolation points $x_k := x_{n,k}$, $-1 \le x_1 < \cdots < x_n \le 1$, as a map of $C[-1,1]$ into itself. Clearly, $\|L_{n-1}\| = \Lambda_n := \max_{-1\le x\le 1}\lambda_n(x)$, where $\lambda_n(x)$ is the Lebesgue function

$$\lambda_n(x) := \sum_{k=1}^n |l_{n,k}(x)|\,.$$

The interpolation points which produce the smallest value Λ_n^* of all Λ_n are not known, (see Notes) but Bernstein [A-1954, vol. 2, p. 107] proves that $\Lambda_n^* = \frac{2}{\pi}\log n + \mathcal{O}(1)$. With Ehlich and Zeller [1966] we shall show that the zeros of the n-th Chebyshev polynomial C_n:

$$x_k := x_{n,k} := \cos t_k\,, \quad t_k := \frac{2k-1}{2n}\pi\,, \qquad k = 1, \ldots, n \tag{2.7}$$

produce the same result asymptotically.

Theorem 2.4. *For the interpolation formula with the points (2.7), the Lebesgue constants Λ_n are*

$$\Lambda_n = \frac{1}{n}\sum_{k=1}^n \cot\frac{(2k-1)\pi}{4n} = \frac{2}{\pi}\log n + \mathcal{O}(1)\,. \tag{2.8}$$

Proof. With the Chebyshev polynomial $C_n(x)$, $x = \cos t$ for $0 \le t \le \pi$, we have (see §6 of Chapter 3)

$$l_{n,k}(x) = \frac{C_n(x)}{(x-x_k)C_n'(x_k)} = \pm\frac{\cos nt}{n(\cos t - \cos t_k)}\sin t_k\,. \tag{2.9}$$

Since

$$\frac{2\sin t_k}{\cos t - \cos t_k} = -\frac{\sin\left(\frac{t+t_k}{2} - \frac{t-t_k}{2}\right)}{\sin\frac{t-t_k}{2}\sin\frac{t+t_k}{2}} = -\cot\frac{t-t_k}{2} + \cot\frac{t+t_k}{2}\,,$$

we have

$$\lambda_n(x) = \sum_1^n |l_{n,k}(x)| = \frac{|\cos nt|}{2n}\sum_{k=1}^n\left|\cot\frac{t-t_k}{2} - \cot\frac{t+t_k}{2}\right|\,.$$

We use the definition $t_k := \frac{(2k-1)\pi}{2n}$ for all $k = 1,\dots,2n$. Since

$$t + t_k = 2\pi + t - t_{2n-k+1}\,,$$

$$\lambda_n(x) \le \mu_n(t) := \frac{|\cos nt|}{2n}\sum_{k=1}^{2n}\left|\cot\frac{t-t_k}{2}\right|\,. \tag{2.10}$$

The function $\mu_n(t)$ is continuous on $\mathbb{T}$ and has period $\frac{\pi}{n}$, for the map $t \to t+\frac{\pi}{n}$ merely permutates the terms of the sum. Thus, μ_n attains equal maxima on the intervals $I_k := [t_k, t_{k+1}]$, $k = 1,\dots,2n$. We examine $\mu_n(t)$ on the interval

$$I_0 := \left[-\frac{\pi}{2n}, \frac{\pi}{2n}\right]\,.$$

On this interval, $\cot\frac{t-t_k}{2}$ is ≤ 0 for $k = 1,\dots,n$, and ≥ 0 for $k = n+1,\dots,2n$. It follows that

$$\mu_n(t) = \frac{\cos nt}{2n}\left(\sum_{k=n+1}^{2n}\cot\frac{t-t_k}{2} - \sum_{k=1}^{n}\cot\frac{t-t_k}{2}\right)\,, \qquad t \in I_0\,. \tag{2.11}$$

We note that $\mu_n'(0) = 0$. We shall show that μ_n' has only one zero on I_0. This will imply that

$$\max_{t\in I_0}\mu_n(t) = \mu_n(0) = \frac{1}{n}\sum_{k=1}^n\cot\frac{(2k-1)\pi}{4n}\,, \tag{2.12}$$

and consequently $\Lambda_n = \lambda_n(1) = \mu_n(0)$. On I_0, μ_n is equal to the polynomial $T \in \mathcal{T}_{n-1}$, given by the trigonometric expression in (2.11). Since

$$\frac{\cos nt}{2n}\cot\frac{t-t_k}{2}\bigg|_{t=t_l} = (-1)^k \quad\text{if}\quad l = k\,, \quad = 0 \quad\text{if}\quad l \ne k\,,$$

the signs of $T(t_k)$ will alternate, except that they will be the same at t_{2n}, t_1 and at t_n, t_{n+1}. Each of the intervals

$$I_k\,, \quad k = 1,\dots,n-1,n+1,\dots,2n-1$$

contains a zero of T. By Rolle's theorem, there are at least $2n-2$ zeros of T' inside $\bigcup_{k=1}^{2n-1} I_k$, hence at most one zero of T' in I_0.

To obtain the asymptotic formula from (2.12), we use the fact that $\cot u - 1/u$ is bounded for $0 < u \le \frac{\pi}{2}$, and so

$$\Lambda_n = \frac{1}{n}\sum_{k=1}^n\frac{4n}{(2k-1)\pi} + \mathcal{O}(1) = \frac{2}{\pi}\sum_{k=1}^n\frac{1}{k-1/2} + \mathcal{O}(1) = \frac{2}{\pi}\log n + \mathcal{O}(1)\,. \qquad \square$$

§ 3. Examples of Linear Polynomial Operators

In §5 we shall see that on $C(\mathbb{T})$, there do not exist operators which preserve trigonometric polynomials of degree n and give an approximation comparable to the best approximation,

$$\|f - U_n f\| \le M E_n(f)\,, \qquad n = 1,2,\dots,$$

with some constant M. But there exist operators which are polynomials of degree $2n-1$ with this property.

Theorem 3.1. *For each positive integer n, there exists a trigonometric polynomial operator V_n of degree $2n-1$ with the properties $V_n(T_n) = T_n$ for each $T_n \in \mathcal{T}_n$ and with*

$$\|f - V_n(f)\| \le 4E_n^*(f)\,, \qquad f \in C(\mathbb{T})\,. \tag{3.1}$$

Proof. To find the operators V_n, we compare the properties of the Fourier sums $s_n(f)$ and the Fejér means $\sigma_n(f)$. The operators s_n leave invariant all trigonometric polynomials T_n of degree $\le n$. The σ_n do not have this property, but as a compensation, they satisfy $\sigma_n(f) \to f$, that is, $\|f - \sigma_n(f)\| \to 0$ for all $f \in C(\mathbb{T})$. To obtain operators with both these properties, we consider the "gliding" average:

$$\frac{1}{p}\left[s_n(f,x) + s_{n+1}(f,x) + \cdots + s_{n+p-1}(f,x)\right]\,. \tag{3.2}$$

If $n = 0$, this is σ_p; if $p = 1$, this is s_n. But we want $p = p(n) \to \infty$. In particular, we choose $p = n$, and obtain the de la Vallée-Poussin sums

$$V_n(f,x) = \frac{1}{n}\left[s_n(f,x) + \cdots + s_{2n-1}(f,x)\right] = 2\sigma_{2n-1}(f,x) - \sigma_{n-1}(f,x)\,. \tag{3.3}$$

We note some simple properties of the operators V_n:

1. V_n is a trigonometric polynomial operator of degree $2n-1$, and $V_n(T_n) = T_n$. This follows from the first relation (3.3).

2. We have $V_n(f) \to f$, for $n \to \infty$, $f \in C$. This follows from the second relation (3.3), since $\sigma_n(f) \to f$.

3. $\|V_n\| \le 3$. Indeed,

$$\|V_n\| \le 2\,\|\sigma_{2n-1}\| + \|\sigma_{n-1}\| = 3\,.$$

We prove (3.1) with the help of the polynomials T_n of best approximation of the function f:

$$\begin{aligned}\|f - V_n(f)\| &= \|f - T_n - V_n(f - T_n)\| \le \|f - T_n\| + \|V_n\| \cdot \|f - T_n\| \\ &\le 4\,\|f - T_n\| = 4E_n(f)\,. \qquad \square\end{aligned}$$

It is easy to generalize Theorem 3.1 somewhat. Let $0 < \varepsilon < 1$ be given. Taking $p = [\varepsilon n]$ in (3.2), we obtain a sequence of polynomial operators V_n' of degrees $\le (1+\varepsilon)n$, for which $\|f - V_n'(f)\| \le \text{const } E_n(f)$.

As an example of an application of Theorem 3.1, we derive:

Theorem 3.2. *For the uniform trigonometric approximation of $g(t) = |\cos t|$ on $\mathbb{T}$,*

$$E_n(g) \sim n^{-1}\,. \tag{3.4}$$

Proof. The Fourier series of g converges. It can easily be found:

$$g(t) = \frac{2}{\pi} + \frac{4}{\pi}\sum_{k=1}^{\infty}(-1)^{k+1}\,\frac{\cos 2kt}{4k^2-1}\,.$$

Since

$$0 = g\left(\frac{\pi}{2}\right) = \frac{2}{\pi} - \frac{4}{\pi}\sum_{k=1}^{\infty}(4k^2-1)^{-1}\,,$$

we obtain for some $M' > 0$,

$$s_m\left(g, \frac{\pi}{2}\right) = \frac{2}{\pi} - \frac{4}{\pi}\sum_{2k\le m}(4k^2-1)^{-1} = \frac{4}{\pi}\sum_{2k>m}(4k^2-1)^{-1} \ge \frac{M'}{m}\,.$$

Therefore,

$$V_n\left(g, \frac{\pi}{2}\right) - g\left(\frac{\pi}{2}\right) = \frac{1}{n}\sum_{m=n}^{2n-1} s_m\left(g, \frac{\pi}{2}\right) \ge \frac{M'}{2n}\,, \qquad n = 1, 2, \dots\,.$$

This proves that $E_n(g) \ge \text{const } n^{-1}$. The opposite inequality $E_n(g) \le \text{const } n^{-1}$ follows from Jackson's Theorem 2.2 of Chapter 7. $\square$

By means of the standard substitution $x = \cos t$ we obtain from this for polynomial approximation on $[-1, +1]$:

Theorem 3.3. $E_n(|x|) \sim n^{-1}$.

Actually, Bernstein [A-1952, vol. 1, p. 172–206] proved that the limit

$$\lim_{n\to\infty}(nE_n(|x|)) = C$$

exists; his computation has been so difficult, that Berstein was able to find only two decimals of C. Using his methods and computers, Varga and Carpenter [1985] found

$$C = 0.2801694990\dots\,.$$

There are no convolutions for functions on $[-1, 1]$. This is one of the reasons why there do not exist formulae for algebraic polynomials which correspond to the Jackson integrals of §2 of Chapter 7. According to Bojanic and DeVore [1969], one can obtain something similar on a part of $[-1, 1]$, if one also uses the extension Theorem 4.1 of Chapter 6.

The smallest positive zeros $\alpha_j := \alpha_{j,n}$, $j = 1, \dots, p$, $\alpha := \alpha_1$ of the Chebyshev polynomial C_n are found from (2.7). They satisfy $0 < \alpha_j \le Cn^{-1}$, with C depending only on p. For simplicity we restrict ourselves to the uniform norm and to moduli of continuity. We define kernels Λ_n and constants λ_n by

$$\Lambda_n(x) = \lambda_n\,\frac{C_n(x)^2}{(x^2-\alpha^2)^2}\,, \qquad \int_{-1}^{+1}\Lambda_n\,dx = 1 \tag{3.5}$$

and the operator $R_n(f)$ by

$$R_n(f, x) = \int_{-\frac{1}{2}}^{\frac{1}{2}} f(t)\Lambda_n(x-t)\,dt\,. \tag{3.6}$$

Proposition 3.4. *For a function $f \in C(I')$, $I' = [-\frac{1}{4}, \frac{1}{4}]$ and for any of its extensions (also denoted by f) onto $[-\frac{1}{2}, \frac{1}{2}]$ with modulus of continuity $\le C\omega(f, h)$, one has $R_n(f) \in C(I')$ and, in the norm of the space $C(I')$,*

$$\|R_n(f) - f\| \le C\left(\omega\left(f, n^{-1}\right) + \|f\|n^{-2}\right)\,. \tag{3.7}$$

Proof. The kernel Λ_n is an even polynomial of degree $2n-4$ and one has

$$\rho_k := \int_{-1}^{+1}|x|^k\Lambda_n\,dx \le \frac{C}{n^k}\,, \qquad k = 0, 1, 2\,. \tag{3.8}$$

For $k = 0$, this is from (3.5). If (3.8) is established for $k = 2$, the statement for $k = 1$ will follow by Cauchy's inequality. To estimate ρ_2, it is very convenient to use the Gaussian quadrature formula (see §8 of Chapter 4) with measure

$$d\mu = (1-x^2)^{-\frac{1}{2}}\,dx\,.$$

For even polynomials P of degree $\le 2n-2$ it has the form

$$\int_{-1}^{+1} P\,d\mu = \frac{\pi}{n}\sum_{j=1}^{n} P(\alpha_j)\,. \tag{3.9}$$

In particular, for $P = x^2\Lambda_n$ we have, using (3.9) twice, both measures dx and $d\mu$, and the inequality $\alpha \le Cn^{-1}$,

$$\begin{aligned}\rho_2 &\le \int_{-1}^{+1} x^2\Lambda_n\,d\mu = \frac{\pi}{n}\,\alpha^2\Lambda_n(\alpha)\\ &\le \frac{C}{n^2}\,\frac{\pi}{n}(1-\alpha^2)\Lambda_n(\alpha) = \frac{C}{n^2}\int_{-1}^{+1}(1-x^2)\Lambda_n\,d\mu \le \frac{C}{n^2}\,.\end{aligned}$$

With the intervals $I := [-1,1]$, $I_x := [-\frac{1}{2}-x, \frac{1}{2}-x]$, for $x \in I'$,

$$R_n(f,x) - f(x) = \int_{I_x}[f(x+u)-f(x)]\Lambda_n(u)\,du - f(x)\int_{I\setminus I_x}\Lambda_n\,du\,.$$

The norm of the second term on the right is

$$\le \|f\|\max\int_{I\setminus I_x}\Lambda_n\,du \le \|f\|\cdot 16\int_{I\setminus I'}u^2\Lambda_n\,du \le C_2\|f\|n^{-2}$$

by (3.8). The norm of the first term does not exceed

$$\int_I \omega(f,u)\Lambda_n(u)\,du \le C_1\omega\left(f,n^{-1}\right)\,,$$

as in (2.3) of Chapter 7. □

From this we obtain Theorem 6.3 of Chapter 7 for $r=1$ and the uniform norm. Indeed, taking $g := f - f(0)$, $P_n := R_n(g) + f(0)$, we have $\|g\| \le \omega\left(f, \frac{1}{4}\right)$ and therefore in $C(I')$

$$\begin{aligned}\|f-P_n\| &= \|g - R_n(g)\|\\ &\le C\left(\omega\left(f,n^{-1}\right) + \omega\left(f,4^{-1}\right)n^{-2}\right)\\ &\le C\omega\left(f,n^{-1}\right)\,.\end{aligned} \tag{3.10}$$

One can also prove the general case of Theorem 6.3 of Chapter 7 for the uniform norm in this way, if one replaces the kernel Λ_n by

$$\Lambda_{n,r}(x) = \lambda_{n,r}\left[\frac{C_n(x)}{(x^2-\alpha_1^2)\cdots(x^2-\alpha_r^2)}\right]^2\,. \tag{3.11}$$

§ 4. Positive Operators

Positive operators have a sufficiently specific structure to deserve separate study. Recall that Theorem 3.1 of Chapter 1 often makes it very easy to check the convergence $U_n f \to f$. Similar to this, there are simple estimates of the norm of the difference $U_n f - f$, given in Theorems 4.2 and 4.3 below. On the other hand, convergence $U_n f \to f$ cannot be too rapid (Theorem 4.1). We encounter here for the first time the phenomenon of *saturation*: there is a certain rapidity of convergence of $\|U_n f - f\|$ to zero, which cannot be overcome by even very strong assumptions about the smoothness or regularity of f. In contrast to this, the partial sums $s_n(f)$ of the Fourier series (not positive operators), admit arbitrary fast (and arbitrary slow) convergence $s_n(f) \to f$.

Our considerations will be for the uniform norm, the first theorem for $A = [-1,1]$.

For a sequence of uniformly bounded linear positive operators U_n on $C(A)$ we put $M := \sup_n \|U_n\|$. We shall use three test functions $\phi_i = x^i$, $i = 0,1,2$, and the quantity

$$\lambda_n := \max_{i=0,1,2}\|\phi_i - U_n(\phi_i)\|\,, \qquad n = 1,2,\ldots\,. \tag{4.1}$$

We also need the functions

$$h(x) = |x|, \quad h_a(x) = |x-a|\,, \qquad h_a^2 = a^2\phi_0 - 2a\phi_1 + \phi_2\,.$$

We have then

$$U_n(h_a^2, a) \le 4\lambda_n\,. \tag{4.2}$$

Indeed,

$$\begin{aligned}0 \le U_n(h_a^2,a) &= U_n(a^2\phi_0 - 2a\phi_1 + \phi_2, a)\\ &= a^2[U_n(\phi_0,a) - \phi_0(a)] - 2a[U_n(\phi_1,a) - \phi_1(a)]\\ &\quad + [U_n(\phi_2,a) - \phi_2(a)]\\ &\le a^2\lambda_n + 2a\lambda_n + \lambda_n \le 4\lambda_n\,.\end{aligned}$$

For a positive operator $U \ge 0$ and functions $f, g \in C[-1,1]$ we have the Cauchy-Schwarz inequality, namely

$$|U(fg,a)| \le U(f^2,a)^{1/2}\,U(g^2,a)^{1/2}\,, \tag{4.3}$$

which is obtained in the usual way, from the relation

$$U\left((f+\lambda g)^2, a\right) \ge 0\,,$$

which is valid for all real λ.

Since

$$U_n(\phi_0, a) \le \|U_n(\phi_0)\| \le M\,,$$

we get from (4.2) and (4.3),

(4.4) $$U_n(h_a, a) \le 2\sqrt{M\lambda_n}\,, \quad n = 1, 2, \dots .$$

Here are a few consequences:

Theorem 4.1 (Korovkin [A-1959]). *If $U_n, n = 1, 2, \dots$, are positive polynomial operators of degree n (so that $U_n(f)$ is a polynomial of degree $\le n$ whenever $f \in C[-1,1]$), then*

(4.5) $$\lambda_n \ne o(n^{-2})\,.$$

In other words, at least one of the functions ϕ_i cannot be approximated by $U_n(\phi_i)$ better than n^{-2}.

Proof. We start with the inequality

$$\big||x| - |a|\big| \le |x - a|\,,$$

that is, with

$$\big|h - |a|\phi_0\big| \le h_a\,.$$

Applying here U_n, we find

$$\begin{aligned}|U_n(h, a) - h(a)| &\le \big|U_n(h - |a|\phi_0, a)\big| + |a|\,|U_n(\phi_0, a) - 1| \\ &\le 2\sqrt{M\lambda_n} + \lambda_n\,.\end{aligned}$$

If $\lambda_n = o(n^{-2})$, this is $o(n^{-1})$. But this contradicts Theorem 3.3 about the degree of approximation of $|x|$ by polynomials. □

There is a similar theorem for $A = \mathbb{T}$, with $\phi_0 = 1$, $\phi_1 = \cos t$, $\phi_2 = \sin t$.

Theorem 4.2. *For a sequence of positive operators U_n and for $f \in C[-1,1]$,*

(4.6) $$\|f - U_n(f)\| \le \|f\|\lambda_n + C_1\omega(f, \sqrt{\lambda_n})\,,$$

where C_1 is a fixed multiple of $M = \sup\|U_n\| \ge 1$. If the U_n reproduce constants, and if $\lambda_n \le 1$, $n = 0, 1, \dots$, then the term $\|f\|\lambda_n$ in (4.6) can be omitted, and C_1 replaced by 3.

Proof. For $a, x \in [-1,1]$ we have

$$|f(x) - f(a)| \le \omega(f, |x-a|) \le \left[1 + \frac{|x-a|}{\sqrt{\lambda_n}}\right]\omega(f, \sqrt{\lambda_n})\,,$$

that is,

$$|f - f(a)\phi_0| \le \left(\phi_0 + h_a/\sqrt{\lambda_n}\right)\omega\left(f, \sqrt{\lambda_n}\right)\,.$$

Since $|U_n(\phi_0, a)| \le M$, we obtain from this and (4.4),

$$\begin{aligned}|U_n(f, a) - f(a)| &\le |f(a)(U_n(\phi_0, a) - \phi_0(a))| + |U_n(f - f(a)\phi_0, a)| \\ &\le \|f\|\lambda_n + \omega\left(f, \sqrt{\lambda_n}\right)\left(M + 2\sqrt{M}\right).\end{aligned}$$

If $U_n(\phi_0) = \phi_0$, then $\|U_n\| = 1$, $M = 1$, and then the first term on the right of the last inequality can be omitted. □

If $f \in C^1[-1,1]$, one can derive in a similar way an estimate of $\|f - U_n f\|$ which contains

$$\sqrt{\lambda_n}\,\omega\left(f', \sqrt{\lambda_n}\right)\,.$$

Both this and Theorem 4.2 were given by Popoviciu for the case of the Bernstein polynomials. We will not derive the the second theorem. We give instead a stronger inequality with $\omega_2(f, \sqrt{\lambda_n})$, whose proof depends on the theory of K-functionals. It will apply to Chebyshev systems on $A = [-1,1]$ or $\mathbb{T}$.

We assume now that ϕ_0, ϕ_1, ϕ_2 is an *extended Chebyshev system* on A. This means that the ϕ_i are twice continuously differentiable and that a nontrivial polynomial in the ϕ_i cannot have more than two zeros, *counting multiplicities*. For instance, 1, $\cos t$, $\sin t$ is an extended Chebyshev system on $\mathbb{T}$. It follows that for each $a \in A$, the Wronskian Δ satisfies

(4.7) $$\Delta(a) := \begin{pmatrix} \phi_0(a), & \phi_1(a), & \phi_2(a) \\ \phi_0'(a), & \phi_1'(a), & \phi_2'(a) \\ \phi_0''(a), & \phi_1''(a), & \phi_2''(a) \end{pmatrix} \ne 0\,.$$

For if $\Delta(a) = 0$, for some $a \in A$, then we could find coefficients c_i, not all zero, for which $P(a) = P'(a) = P''(a) = 0$ for the polynomial

$$P = c_0\phi_0 + c_1\phi_1 + c_2\phi_2\,.$$

For a fixed $a \in A$, it is convenient to use the linear combinations

$$\psi_j(x) := \psi_j(x, a)\,, \quad j = 0, 1, 2\,,$$

of the ϕ_i which satisfy

(4.8) $$\psi_j^{(i)}(a) = \delta_{i,j}\,, \quad i, j = 0, 1, 2\,.$$

The coefficients $c_{i,j} = c_{i,j}(a)$ of the linear combinations

$$\psi_j = \sum_{i=0}^{2} c_{i,j}\phi_i$$

can be found from systems of linear equations with the determinant $\Delta(a)$. It follows that the $c_{i,j}$ are *uniformly bounded for $a \in A$*. Let the λ_n be defined by (4.1) for the φ_i, and let $C, C_1, \dots$ stand for constants which depend on the system ϕ_i but not on a. We see that

(4.9) $$\max_{j=0,1,2} \|\psi_j - U_n(\psi_j)\| \le C_1 \lambda_n .$$

We also note that for some $C_2 > 0$,

(4.10) $$\psi_2(x) = \psi_2(x,a) \ge C_2(x-a)^2 , \qquad a, x \in A .$$

This follows from the expansion

$$\psi_2(x) = \frac{1}{2}(x-a)^2[1+o(1)] ,$$

where $o(1)$ is uniformly small (in the variable a) for $x \to a$, and the fact that $\psi_2(x) \ne 0$ for $x \ne a$.

Theorem 4.3 (Freud [1968]). *For each extended Chebyshev system ϕ_0, ϕ_1, ϕ_2 on A there is a constant C so that for each sequence of positive operators U_n and each $f \in C(A)$,*

(4.11) $$\|f - U_n(f)\| \le C \left\{ \omega_2 \left(f, \sqrt{\lambda_n} \right) + \|f\| \lambda_n \right\} , \qquad n = 1, 2, \ldots .$$

Proof. First let $f \in C^2(A)$. We would like to estimate $|f - U_n f|$. Selecting $a \in A$, we define

$$P(x) := P(f,a,x) = f(a)\psi_0(x) + f'(a)\psi_1(x) .$$

Then $P(a) = f(a)$, $P'(a) = f'(a)$. If $L = \|f\| + \|f''\|$, then by Theorem 5.6 of Chapter 2, $\|f'\| \le$ const L. We apply U_n to P and get, using (4.9),

(4.12) $$|f(a) - U_n(P,a)| \le \|P - U_n(P)\| \le C_2 L \lambda_n .$$

Next, for some $\xi \in A$,

$$|f(x) - P(x)| = |f''(\xi) - P''(\xi)| \, \frac{(x-a)^2}{2} \le C_3 L (x-a)^2 \le C_4 L \psi_2(x) .$$

Applying U_n to this inequality,

(4.13) $$\begin{aligned} |U_n(f,a) - U_n(P,a)| &\le C_4 L |U_n(\psi_2, a)| \\ &= C_4 L |U_n(\psi_2,a) - \psi_2(a)| \le C_5 L \lambda_n . \end{aligned}$$

Adding (4.12) and (4.13) we obtain, because $a \in A$ was arbitrary,

(4.14) $$\|f - U_n(f)\| \le C \left(\|f\| + \|f''\| \right) \lambda_n .$$

We now use Theorem 5.2 of Chapter 7 to derive (4.11) from this relation. □

Remark. As a special case of (4.11) we obtain

(4.15) $$\|f - U_n(f)\| \le C \left\{ \sqrt{\lambda_n}\, \omega \left(f', \sqrt{\lambda_n} \right) + \|f\| \lambda_n \right\}$$

and also a form of (4.6). However, the earlier proof of (4.6) has been essentially simpler.

In some special cases, one can improve (4.11), to *local estimates*, that is, to estimates of $|f(x) - U_n(f,x)|$ with an upper bound *depending on x*.

We assume that

$$\phi_0 = 1 , \quad \phi_1 = x , \quad \phi_2 = x^2 , \quad A = [-1,1] ,$$

then

$$\psi_0 = 1 , \quad \psi_1 = (x-a) , \quad \psi_2 = (x-a)^2 .$$

We further assume that the U_n reproduce ϕ_i, $i = 0, 1$. Then:

(4.16) *For any convex function g, $U_n(g,x) \ge g(x)$ for all x.*

To see this, let $-1 \le x_0 \le 1$ be arbitrary and let $y = \alpha x + \beta$ be the supporting line to $y = g(x)$ at x_0. Then $g(x) \ge \alpha x + \beta$, and therefore

$$U_n(g,x) \ge \alpha x + \beta .$$

Putting here $x = x_0$, we obtain the inequality corresponding to (4.14) at x_0.

Theorem 4.4. *For positive operators U_n which reproduce linear functions, with $\lambda_n(x) := U_n(\phi_2, x) - \phi_2(x)$,*

(4.17) $$|f(x) - U_n(f,x)| \le C \omega_2 \left(f, \lambda_n(x) \right) .$$

The *proof* is a simplification of that of Theorem 4.3. The polynomial P is now

$$P(x) = f(a) + f'(a)(x-a) , \quad U_n(P,a) = f(a) , \quad \text{and}$$
$$|f(x) - P(x)| \le \frac{1}{2} \|f''\| (x-a)^2 .$$

Since $U_n(\psi_2, a) = U_n(\phi_2, a) - a^2 = \lambda_n(a)$, an application of U_n to the last inequality yields

$$|U_n(f,a) - f(a)| \le \frac{1}{2} \|f''\| \lambda_n(a) .$$

We appeal to Theorem 5.3 of Chapter 7 to obtain (4.17). □

§ 5. Projections onto Spaces Spanned by Exponentials

The last three sections of this chapter will be devoted to the study of projections onto finite dimensional spaces, in particular, projections with the smallest norm. The simplest projection in the space $L_1(\mathbb{T})$ is the Fourier series sum $s_n(f)$. We shall see that it has certain minimal properties.

But this can be generalized. Let D be any finite set of distinct integers. Let X_D be the space spanned by the exponentials e^{ikx}, $k \in D$. If D is symmetric, that is, if $k \in D$ implies $-k \in D$, X_D is also spanned by $\cos kx,\ \sin kx,\ k \in D$. The *Fourier projection* S_D of a function $f \in L_1(\mathbb{T})$ onto X_D is the sum

$$S_D(f,x) := \sum_{k \in D} c_k\, e^{ikx}\,, \quad c_k = \widehat{f}(k) = \frac{1}{2\pi}\int_{\mathbb{T}} f(t) e^{-ikt}\, dt\,. \tag{5.1}$$

In this and the next section the functions will be *complex valued*. By Theorem 4.2 of Chapter 2, the norm of S_D in $C(\mathbb{T})$ and in $L_1(\mathbb{T})$ is precisely the integral

$$\|S_D\| = \int_{\mathbb{T}} \left| \sum_{k \in D} e^{ikx} \right| dx\,. \tag{5.2}$$

The main theorem of this section is that S_D has smallest norm among all projections of $C(\mathbb{T})$ onto X_D.

By f_a we denote the *a-translation* of a function $f \in C(\mathbb{T})$, $f_a(x) := f(x+a)$. Then $\|f_a\|_\infty = \|f\|$, and $(f+g)_a = f_a + g_a$. It follows that the operation $f \to f_a$ is a continuous function of a.

We begin by establishing a useful formula:

Theorem 5.1. *For each projection P of $C(\mathbb{T})$ onto X_D, one has*

$$\|P\| \ge \|S_D\|\,, \tag{5.3}$$

$$\frac{1}{2\pi}\int_{\mathbb{T}} P(f_t, x-t)\, dt = S_D(f,x)\,. \tag{5.4}$$

Proof. Formula (5.4) is due to Berman [1952] and Marcinkiewicz. We show first that the integrand is a continuous function of $x, t \in \mathbb{T}^2$. Let x_0, t_0 be fixed, then

$$\begin{aligned}|P(f_t,x) - P(f_{t_0},x_0)| \le{}& |P(f_{t_0},x) - P(f_{t_0},x_0)| \\ &+ |P(f_t,x) - P(f_{t_0},x)|\,.\end{aligned}$$

The first term on the right is small if x is close to x_0, because $P(f_{t_0},x)$ is a continuous function of x. The second term is small if t is close to t_0 because it does not exceed $\|P\| \cdot \|f_t - f_{t_0}\|$.

Thus, the integral (5.4) exists. We denote it by $A(f,x)$ and prove that $A(f) = S_D(f)$ for all f. Both operators S_D and A are linear and bounded. For A this follows from

$$\begin{aligned}|A(f,x)| &\le \frac{1}{2\pi}\int_{\mathbb{T}} |P(f_t, x-t)|\, dt \\ &\le \frac{1}{2\pi}\int_{\mathbb{T}} \|P\| \cdot \|f_t\|\, dt = \|P\| \cdot \|f\|\,, \quad x \in \mathbb{T}\,.\end{aligned}$$

It follows that

$$\|A\| \le \|P\|\,. \tag{5.5}$$

The equality (5.4), that is $A(f) = S_D(f)$, holds for the following functions:

(a) $f(x) = e^{ikx}$, $k \in D$. Then $f_t(x) = e^{ikt}\, e^{ikx}$, and since P is a projection onto X_D,

$$P(f_t, x) = f_t(x)\,, \quad P(f_t, x-t) = f(x)\,,$$

and (5.4) follows.

(b) $f(x) = e^{ilx}$, $l \notin D$. Then

$$P(f_t, x-t) = e^{ilt}\, P(f, x-t)\,.$$

Here $P(f, x-t)$ is a linear combination (with coefficients depending on x) of exponentials e^{ikt}, $k \in D$, which are orthogonal to e^{ilt}. Therefore, again

$$A(f,x) = \frac{1}{2\pi}\int_{\mathbb{T}} e^{ilt}\, P(f, x-t)\, dt = 0 = S_D(f,x)\,.$$

From (a) and (b) it follows that $A(f) = S_D(f)$ for all trigonometric polynomials f, and since they are dense in $C(\mathbb{T})$, the operators A and S_D coincide. □

One can draw many interesting conclusions from Theorem 5.1. The simplest is:

Corollary 5.2. *For a projection P of $C(\mathbb{T})$ onto the trigonometric polynomials $\mathcal{T}_n$ of degree $\le n$,*

$$\|P\| \ge \|s_n\| \ge \text{const}\ \log n\,. \tag{5.6}$$

Theorem 5.3. *For each $n = 1, 2, \dots$, let U_n be a bounded linear operator which maps $C(\mathbb{T})$ into $\mathcal{T}_n$. There cannot exist a continuous function*

$$\phi(u) \ge 0\,, \quad u \ge 0\,, \quad \phi(0) = 0$$

for which

$$\|f - U_n(f)\| \le \phi(E_n(f))\,, \quad f \in C(\mathbb{T})\,, \quad n = 1, 2, \dots\,. \tag{5.7}$$

Proof. It follows from (5.7) that $U_n(T) = T$ for $T \in \mathcal{T}_n$, so that the U_n are projections. Again from (5.7), $U_n(f) \to f$ as $n \to \infty$ for $f \in C(\mathbb{T})$. The norms $\|U_n\|$ must be bounded, and this contradicts (5.6). □

We need an analogue of Theorem 5.1 which gives lower bounds for the norms of projections onto the space of algebraic polynomials. Let D be a finite set of nonnegative integers and let P be a projection of the subspace of even functions in $C(\mathbb{T})$ onto the even trigonometric polynomials $\sum_{k \in D} c_k \cos kx$. We can extend P to all of $C(\mathbb{T})$ by writing $P^*(f) = P(F)$ where $F(x) := \frac{1}{2}(f(x) + f(-x))$ is the even part of f. Obviously, P^* is linear and $\|P^*\| = \|P\|$.

One way to find such a projection P is to take $D^* := D \cup (-D)$; then S_{D^*} is such a P. We note that $S^*_{D^*} \ne S_{D^*}$. The analogue of Theorem 5.1 is that for any projection P of the above type, we have

(5.8) $$\frac{1}{2\pi}\int_{\mathbb{T}} P^*(f_t+f_{-t},x-t)\,dt = S^*_{D^*}(f,x)$$

(5.9) $$\|P\| = \|P^*\| \ge \frac{1}{2}\|S^*_{D^*}\| = \frac{1}{2}\|S_{D^*}\|\,.$$

The proof is the same as for Theorem 5.1. In (a), $f(x) = e^{ikx}$, $k \in D^*$, and $f_{\pm t}(x) = e^{ik(x\pm t)}$, so that $P^*(f_t+f_{-t}) = 2\cos kx\cos kt$. Therefore, the integrand in (5.8) is $2\cos kt\cos k(x-t)$, and the left hand side is $\frac{1}{\pi}\int_{\mathbb{T}}\cos kx\cos^2 kt\,dt = \cos kx$ which is also the right side of (5.8). If $f(x) = e^{ilx}$, $l \notin D^*$, then the integrand is $T(x-t)\cos lt$ where T is a cosine polynomial with frequencies from D. Since $l \notin D$, the left side of (5.8) is zero and so is the right.

If D consists of all k, $0 \le k \le n$, we have $\|P\| \ge \frac{1}{2}\|s_n\|$.

Corollary 5.4. *If U is a projection of $C[-1,1]$ onto $\mathcal{P}_n$, then*

(5.10) $$\|U\| \ge \frac{1}{2}\|s_n\| = \frac{2}{\pi^2}\log n + \mathcal{O}(1)\,.$$

Proof. The map $V : f(x) \to f(\cos t)$ is an isometry of $C[-1,1]$ into $C(\mathbb{T})$. Moreover, $\|V\| = 1$, and $P := VUV^{-1}$ is a projection of $C[0,\pi]$ onto the even trigonometric polynomials. We have $\|U\| = \|P\|$. □

We shall finally mention that the *lacunary* sets of exponentials, e^{in_kx}, $k = 1,\dots,N$, for which $n_{k+1}/n_k \ge \lambda$ for some fixed $\lambda > 1$, behave in a very different way.

Theorem 5.5. *For a lacunary set of exponentials e^{in_kx}, $k = 1,\dots,N$, with $n_{k+1}/n_k \ge \lambda > 1$ one has*

(5.11) $$\|S_D\| \ge C\sqrt{N}\,,$$

with C depending only on λ.

Proof. The number $a(s)$ of representations $s = n_k + n_l$, $k,l \le N$, for integers $s > 0$ does not exceed a constant which depends only on λ. Indeed, if $n_l \le n_k$, we must have $s/2 \le n_k \le s$, and there are at most $1 + \frac{\log 2}{\log\lambda}$ such n_k.

By Hölder's inequality, with $p = 3$,

(5.12) $$N = \int_T\left|\sum_{k=1}^N e^{in_kt}\right|^2 dt \le \left\{\int_T\left|\sum e^{in_kt}\right|^4 dt\right\}^{1/3}\left\{\int_T\left|\sum e^{in_kt}\right|dt\right\}^{2/3}$$

The second factor on the right is $\|S_D\|^{2/3}$. For the first factor,

$$\int_T\left|\sum\right|^4 dt = \int_T\left|\left(\sum e^{in_kt}\right)^2\right|^2 dt = \int_T\left|\sum a(s)e^{ist}\right|^2 dt\,.$$

There are at most N^2 terms in the last sum. Thus, the integral does not exceed const N^2, and we obtain (5.11) from (5.12). □

§ 6. Lower Bounds

The main result of this section is the following fact:

Theorem 6.1. *The projection of $C(\mathbb{T})$ onto the space Y spanned by N exponentials e^{in_kt}, $k = 1,\dots,N$, where n_k are integers, has norm*

(6.1) $$\ge C\log N$$

with a constant C, independent of the n_k.

From §5 we know that the Fourier projection onto Y has the smallest norm among all projections, and by Theorem 1.2, its norm is the integral (6.2) below. Therefore, Theorem 6.1 follows from

Theorem 6.2. *For some absolute constant $C > 0$,*

(6.2) $$\int_{\mathbb{T}}\left|\sum_{k=1}^N e^{in_kt}\right| dt \ge C\log N\,.$$

It may appear intuitively obvious that the exponentials e^{ikt}, $k = 0,\pm1,\dots$, if they are N in number, "are the best possible", and that they give rise to a projection of the smallest norm. However, as far as we know, this statement has not been proved.

The inequality (6.2) is known as the *Littlewood conjecture*. After many attempts to its proof, which have resulted in weaker statements, Theorem 6.2 has been established only in 1981 independently by McGehee, Pigno and B. Smith [1981] and Konjagin [1981]. Both proofs are of considerable subtlety. We follow the first paper, because it gives also an interesting intermediate result, Theorem 6.3.

For functions $f \in L_1(\mathbb{T})$ we denote by $\sum_{-\infty}^{+\infty}\widehat{f}(k)e^{ikt}$ their complex Fourier series; the *spectrum* of f is the set of k with $\widehat{f}(k) \ne 0$. Of particular importance are functions with a positive spectrum, for which $\widehat{f}(k) = 0$ for $k < 0$. They form the *Hardy class* $H_1(\mathbb{T})$. For $f \in H_1(\mathbb{T})$, the function $\sum_{-\infty}^{+\infty}\widehat{f}(k)z^k$ is analytic in $|z| < 1$, and has boundary values $f(t)$ for $z = e^{it}$ for radial approach within the circle (see Duren [B-1970, Ch.1]). The subspace $H_\infty(\mathbb{T})$ consists of bounded functions on $\mathbb{T}$; they are identical with boundary values of bounded analytic functions in $|z| < 1$. It follows from this that $h \in H_\infty(\mathbb{T})$ implies $\Phi(h) \in H_\infty(\mathbb{T})$, if Φ is an entire function and if $\Phi(h(t))$ is bounded.

Theorem 6.3 (McGehee, Pigno and Smith). *If a function $f \in L_1(\mathbb{T})$ has a spectrum contained in $S = \{n_1 < n_2 < \cdots\}$, then*

$$\sum \frac{|\widehat{f}(n_k)|}{k} \le 30\|f\|_1 . \tag{6.3}$$

We begin the proof of Theorem 6.3 by writing the sequence S as $\cup_{j=0}^{\infty} S_j$, with subsequences S_j of length 4^j. Thus, $S_0 := \{n_1\}$ and S_j, $j \ge 1$, consist of the terms n_k for which $2+4+\cdots+4^{j-1} \le k \le 1+4+\cdots+4^j$. For $n_k \in S_j$,

$$3k \ge 3(2+4+\cdots+4^{j-1}) \ge 1+3\frac{4^j-1}{4-1} = 4^j . \tag{6.4}$$

We define trigonometric polynomials T_j, $j=1,2,\ldots$, whose spectra are S_j by means of

$$|\widehat{T}_j(n)| = 4^{-j}, \ n \in S_j , \qquad \widehat{T}_j(n)\widehat{f}(n) \ge 0, \ n = 0, \ \pm 1, \ldots . \tag{6.5}$$

Then obviously

$$\|T_j\|_2 = 2^{-j} , \qquad \|T_j\|_\infty \le 1 . \tag{6.6}$$

We need the Fourier series of the $|T_j|$

$$|T_j(t)| = \sum_{-\infty}^{+\infty} c_k^{(j)} e^{ikt} ,$$

and a general remark. If a function $g \in L_2$ with the Fourier series $\sum_{-\infty}^{+\infty} c_k e^{ikt}$ has real values, then

$$g(t) = Re\, h(t) , \quad h(t) = c_0 + 2\sum_{-\infty}^{-1} c_k e^{ikt} . \tag{6.7}$$

Indeed, in this case we have $\bar{c}_k = c_{-k}$, $k = 0, \pm 1, \ldots$.

Let h_j, $j = 1,2,\ldots$, be functions derived in this way from $|T_j|$. They have the following properties:

$$\begin{aligned} \|h_j\|_2 &= \sqrt{|c_0|^2 + 4\sum_{-\infty}^{-1} |c_n|^2} \le \sqrt{2}\,\sqrt{\sum_{-\infty}^{+\infty} |c_n|^2} \\ &= \sqrt{2}\,\|T_j\|_2 < \tfrac{3}{2}2^{-j} , \qquad j = 1,2,\ldots . \end{aligned} \tag{6.8}$$

Also, $Re\, h_j(t) \ge 0$, and therefore e^{-h_j} is bounded: $|e^{-h_j}| \le e^{-|T_j|}$. By the remark preceding Theorem 6.3, the spectrum of e^{-h_j} is negative.

The idea of the proof is to replace the T_j by new functions F_m which are bounded like the T_j and whose Fourier coefficients $\widehat{F}_m(n)$ behave like those of $\widehat{T}_j(n)$ for many values of j. We define inductively

$$F_0 = \tfrac{1}{5}T_0 , \quad F_{j+1} = F_j e^{-h_{j+1}/4} + \tfrac{1}{5}T_j , \qquad j = 0,1,\ldots . \tag{6.9}$$

Lemma 6.4. *The functions F_j have the properties*

$$\|F_j\|_\infty \le 1 \tag{6.10}$$

$$\widehat{F}_m(n) = \tfrac{1}{5}(1+\delta)\widehat{T}_j(n) , \quad n \in S_j , \quad j = 0,\ldots,m \tag{6.11}$$

where δ (which depends on n) satisfies $|\delta| \le \frac{1}{2}$.

Proof. We shall use the inequality

$$e^{-\frac{x}{4}} + \tfrac{x}{5} \le 1 , \qquad 0 \le x \le 1 . \tag{6.12}$$

Indeed, the function $\phi(x) = e^{-x/4} + \frac{x}{5} - 1$ has value 0 at $x = 0$, ϕ' changes sign on $[0,1]$, first it is < 0, then > 0. Hence it is sufficient to check that $\phi(1) < 0$, that is, that $e^{-1/4} < \frac{4}{5}$ or $e > (\frac{5}{4})^4$.

Therefore, by induction,

$$|F_{j+1}(t)| \le e^{-\frac{1}{4}Re\, h_{j+1}} + \tfrac{1}{5}|T_{j+1}| \le e^{-\frac{1}{4}|T_{j+1}(t)|} + \tfrac{1}{5}|T_{j+1}(t)| \le 1 .$$

To establish (6.11), we show that

$$|\widehat{F}_m(n) - \tfrac{1}{5}\widehat{T}_j(n)| \le \tfrac{1}{10}\widehat{T}_j(n) , \qquad n \in S_j , \qquad j = 0,\ldots,m . \tag{6.13}$$

From the definition (6.9),

$$F_m = \frac{1}{5}\sum_{l=0}^{m-1} T_l e^{-\frac{1}{4}(h_{l+1}+\cdots+h_m)} + \frac{1}{5}T_m . \tag{6.14}$$

The Fourier series of F_m can be obtained by multiplying out and adding the Fourier series of the different functions in the sum (6.14). For $l < j$, the spectrum of T_l is to the left of S_j, and the spectrum of $e^{-\frac{1}{4}(h_{l+1}+\cdots+h_m)}$ is negative. Hence the spectrum of $T_l\ e^{-\frac{1}{4}(h_{l+1}+\cdots+h_m)}$ is to the left of S_j. These terms can be omitted from the sum (6.14), without changing $\widehat{F}_m(n)$. In particular, this shows that $\widehat{F}_m(n) = \frac{1}{5}\widehat{T}_m(n)$, $n \in S_m$, so we can further restrict our attention to the case $j < m$ in the remainder of the proof. We can also add on the right in (6.14) any linear combination of T_k, $k \ne j$. Hence $\widehat{F}_m(n) = \widehat{G}(n)$, $n \in S_j$, $0 \le j \le m$, for the function

$$G = \frac{1}{5}T_j + \frac{1}{5}\sum_{l=j}^{m-1} T_l \left(e^{-\frac{1}{4}(h_{l+1}+\cdots+h_m)} - 1\right) .$$

By the Cauchy-Schwarz inequality,

$$\left|\widehat{F}_m(n) - \frac{1}{5}\widehat{T}_j(n)\right| \le \frac{1}{5}\sum_{l=j}^{m-1} \|T_l\|_2\ \|e^{-\frac{1}{4}(h_{l+1}+\cdots+h_m)} - 1\|_2 . \tag{6.15}$$

We need here the inequality

$$\|e^{-h} - 1\|_2 \le \|h\|_2 , \tag{6.16}$$

which is valid for a function $h \in L_\infty(\mathbb{T})$ with $Re\, h \geq 0$. This can be seen from

$$|1 - e^{-z}| = |\int_0^z e^{-t}\, dt| \leq |z|\,, \qquad Re\, z \geq 0\,.$$

Therefore, (6.15), (6.6) and (6.8) imply that for $n \in S_j$, $0 \leq j < m$,

$$\left|\widehat{F}_m(n) - \frac{1}{5}\widehat{T}_j(n)\right| \leq \frac{1}{5}\sum_{l=j}^{m-1} 2^{-l}\, \frac{3}{8}(2^{-l-1} + \cdots + 2^{-m})$$
$$\leq \frac{3}{40}\sum_{l=j}^{\infty} 2^{-l}2^{-l} = \frac{1}{10}4^{-j} = \frac{1}{10}|\widehat{T}_j(n)|\,,$$

and this implies (6.11). □

We can now complete the proof of Theorem 6.3. Let m be fixed, let $n = n_k \in S_j$, $0 \leq j \leq m$. Since $\widehat{T}_j(n)\widehat{f}(n) \geq 0$, for all n

$$\widehat{F}_m(n)\widehat{f}(n) = \tfrac{1}{5}(1+\delta)\widehat{T}_j(n)\widehat{f}(n) = \tfrac{1}{5}(1+\delta)4^{-j}|\widehat{f}(n)|\,,$$

$$Re\,\{\widehat{F}_m(n)\widehat{f}(n)\} \geq \tfrac{1}{10}4^{-j}|\widehat{f}(n)|\,.$$

From (6.4), $4^j < 3k$, therefore with $B_m = \cup_{j\leq m} S_j$,

$$Re \sum_{n\in B_m} \widehat{F}_m(n)\widehat{f}(n) \geq \frac{1}{30}\sum_{n_k\in B_m}\frac{|\widehat{f}(n_k)|}{k}\,.$$

On the other hand, since $\|F_m\|_\infty \leq 1$,

$$Re \sum_{n\in B_m} \widehat{F}_m(n)\widehat{f}(n) = Re\int_T F_m f\, dt \leq \|f\|_1\,.$$

Letting $n \to \infty$, we arrive at (6.3). □

§ 7. Projections in Arbitrary Banach Spaces

For a real Banach space X and any of its n-dimensional subspaces Y, we are interested in projections of X onto Y whose norm is small. It is fairly easy to show the existence of a projection of norm $\leq n$. But this bound is not the best possible. The striking result of Kadec and Snobar is that there is always a projection of norm $\leq \sqrt{n}$ (see Theorem 7.6 below). If X is a Hilbert space, there is a projection of norm 1 – the orthogonal projection of X onto Y. The spaces L_p, $1 \leq p \leq +\infty$, take an intermediate position, with projections of norm $n^{|(1/2)-(1/p)|}$ (see Theorem 7.5). This is a result of Lewis [1978], but we follow the basic ideas of the proof of Lorentz and Tomczak-Jaegermann [1984].

In what follows, X will be either the L_p space, $1 < p < +\infty$ for a σ-finite measure space $(A, \mathcal{A}, \mu)$, or the space $C(A)$ of continuous functions on a compact Hausdorff space A. The space L_1 will be treated later.

One of our tools will be the spaces X^n of n-tuples of elements of X. We use the generic notations $\mathbf{h} = (h_1, \ldots, h_n)$, $h_j \in X$, $j = 1, \ldots, n$, and

$$H(x) := \sqrt{\sum\nolimits_1^n h_j(x)^2}\,. \tag{7.1}$$

Then X^n is a Banach space with the norm

$$\|\mathbf{h}\| := \|\mathbf{h}\|_{X^n} := \|H\|_p\,, \qquad 1 < p \leq \infty\,. \tag{7.2}$$

Since each $\mathbf{h}$ is a sum of n tuples with only one of coordinates $\neq 0$, we obtain the representation for a bounded linear functional Ψ on X^n:

$$\Psi(\mathbf{h}) = \sum_{j=1}^{n} \Psi_j(h_j)\,,$$

with bounded linear functionals Ψ_j on X. For $1 \leq p < \infty$, we have $\Psi_j(h) = \int_A h g_j\, d\mu$, $g_j \in L_{p'}$, so that

$$\Psi(\mathbf{h}) = \int_A \sum_{j=1}^{n} h_j g_j\, d\mu\,. \tag{7.3}$$

If $p = \infty$, $X = C(A)$, then $\Psi_j(h) = \int_A h\, d\nu_j$ with finite signed Borel measures ν_j on A. If μ_j is the total variation of ν_j, then $\mu = \sum_1^n \mu_j$ is a finite Borel measure, and by the Radon-Nikodym theorem, $d\nu_j = g_j\, d\mu$, $j = 1, \ldots, n$, where g_j, with $|g_j| \leq 1$, is a μ-measurable function on A. (The introduction of the measure $d\mu$, needed for $p = \infty$, will allow us to treat all cases $1 < p \leq \infty$ simultaneously.) Thus we have (7.3) for all $1 < p \leq +\infty$; for $p = \infty$, however, the measure $d\mu = \sum_1^n d\nu_j$ depends on the functional Ψ. For $p = \infty$, we fix Ψ and define $\|G\|_1 : = \int_A G\, d\mu$.

Lemma 7.1. *The norm of a functional* (7.3) *on* X^n *is given by*

$$\|\Psi\| = \|G\|_{p'}\,, \qquad 1 \leq p \leq +\infty\,. \tag{7.4}$$

Proof. First let $1 \leq p < +\infty$. For $\mathbf{h} \in X^n$,

$$|\Psi(\mathbf{h})| \leq \int_A HG\, d\mu \leq \|H\|_p\, \|G\|_{p'} = \|G\|_{p'}\, \|\mathbf{h}\|\,.$$

On the other hand, for given $\mathbf{g}$, we select an $h \geq 0$ with the properties $\|h\|_p = 1$, $h(x) = 0$ if $G(x) = 0$, and $\int_A Gh\, d\mu = \|G\|_{p'}$. For the functions $h_j(x) : = \frac{g_j(x)}{G(x)}h(x)$, if $G(x) \neq 0$, $h_j(x) := 0$ otherwise, $j = 1, \ldots, n$, we have $H = h$, $\|\mathbf{h}\| = 1$ and $\Psi(\mathbf{h}) = \|G\|_{p'}$. This proves the assertion. For $p = \infty$, the proof is similar. □

Now let Y be an n-dimensional subspace of X; then Y^n is a subspace of X^n. We fix a set $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)$ of n linearly independent linear functionals $\phi_j \in Y^*$, $j = 1, 2, \dots, n$, and discuss the problem: find $\mathbf{f} := (f_1, \dots, f_n) \in Y^n$ so that

$$\det\big[\phi_i(f_j)\big]_{i,j=1}^{n} = \max_{\|H\| \le n^{1/p}} \det[\phi_i(h_j)] = : D\,, \tag{7.5}$$

where $\mathbf{h}$ is an element of Y^n subject to the condition $\|H\|_p \le n^{1/p}$. Since the unit ball in Y is compact, extremal solutions $\mathbf{f}$ to (7.5) exist. We call each extremal set $\mathbf{f} = (f_1, \dots, f_n)$ a *maximal basis* for Y. The functions f_i must be linearly independent. It follows that $D > 0$. Moreover, for the extremal solution we have $\|F\|_p = n^{1/p}$, for otherwise, replacing the f_j, $j = 1, \dots, n$ by their suitable multiples, we obtain a larger determinant D.

If $\boldsymbol{\psi} = (\psi_1, \dots, \psi_n)$ is any other basis for Y^*, then $\mathbf{f}$ is maximal also for $\boldsymbol{\psi}$. Indeed, the $\boldsymbol{\psi}_j$ are expressible in terms of the ϕ_i by means of a linear transformation with a nonsingular matrix T. To pass from the ϕ_i to the ψ_i in (7.5) we have merely to multiply both sides by $\det T$.

Lemma 7.2. *Let $X = L_p$ if $1 < p < \infty$, $X = C$ if $p = \infty$, and let Y^n be one of its n-dimensional subspaces. If $\mathbf{f}$ is a maximal basis for Y and $\boldsymbol{\phi} = (\phi_1, \dots, \phi_n)$ is its dual basis for Y^* defined by*

$$\phi_i(f_j) = \delta_{i,j}\,, \qquad i, j = 1, \dots, n\,, \tag{7.6}$$

then

$$\Phi(\mathbf{h}) = \sum_{i=1}^{n} \phi_i(h_i) \tag{7.7}$$

is a linear functional on Y^n with norm $n^{1/p'}$.

Proof. With the choice (7.6), we have $D = 1$ in (7.5). Since $\|\mathbf{f}\| = \|F\|_p = n^{1/p}$, and $n = \sum \phi_i(f_i) = \Phi(\mathbf{f})$, it follows from (7.4) that $\|\Phi\| \ge n^{1/p'}$. To see that $\|\Phi\| \le n^{1/p'}$, we let $\mathbf{h} = (h_1, \dots, h_n)$ be an arbitrary element of Y^n. With the identity matrix $I = [\phi_i(f_j)]$ and $A : = [\phi_i(h_j)]$, we form for small $t > 0$:

$$I + tA = [\phi_i(f_j + th_j)]\,.$$

For a matrix $A = [a_{i,j}]$, the derivative of the polynomial in t, $\det(I + tA)$ at $t = 0$ is the *trace* $\operatorname{tr} A = a_{11} + \dots + a_{nn}$. Therefore, we have the expansion

$$\det(I + tA) = 1 + t \operatorname{tr} A + \dots = 1 + t\Phi(\mathbf{h}) + \mathcal{O}(t^2)\,. \tag{7.8}$$

On the other hand, let

$$c := c(t) := n^{-1/p}(\|F\|_p + t\|H\|_p) = 1 + tn^{-1/p}\|H\|_p\,,$$

then for $f_j^* = c^{-1}(f_j + th_j)$, $j = 1, \dots, n$, we have $F^* \le c^{-1}(F + tH)$ and consequently, $\|F^*\|_p \le n^{1/p}$.

Hence, by the maximality of $\mathbf{f}$,

$$\det(I + tA) = c^n \det\big[\phi_i(f_j^*)\big] \le c^n = 1 + tn^{1/p'}\|H\| + \mathcal{O}(t^2)\,.$$

Comparing this with (7.8) for small t, we see that

$$\Phi(\mathbf{h}) \le n^{1/p'}\|H\|_p = n^{1/p'}\|\mathbf{h}\|\,. \qquad \square$$

We would like to extend the functional (7.7) from Y^n to X^n, without increasing its norm. By the Hahn-Banach theorem, such extensions Ψ exist; they are of the form (7.3), and have the norm

$$n^{1/p'} = \|\Phi\| = \|\Psi\| = \|G\|_{p'}\,. \tag{7.9}$$

The relation $\Psi(\mathbf{h}) = \Phi(\mathbf{h})$ for $\mathbf{h}$ with $h_j = 0$, $j \ne i$, and $h_i = h \in Y$ yields

$$\phi_i(h) = \int_A g_i h \, d\mu\,. \tag{7.10}$$

This means that the functionals $\phi_i \in Y^*$ of the dual basis Y^* can be identified with elements $g_i \in L_{p'}$.

With the functions $\mathbf{f}, \mathbf{g}$ we get

Proposition 7.3. *For an n-dimensional subset $Y \subset X = L_p$, $1 < p \le +\infty$, there exist functions $f_j \in Y$, $g_j \in L_{p'}$, $j = 1, \dots, n$, for which $\|F\|_p = n^{1/p}$, $\|G\|_{p'} = n^{1/p'}$,*

$$\int_A f_i g_j \, d\mu = \delta_{i,j}\,, \qquad i, j = 1, \dots, n\,, \tag{7.11}$$

and, in addition, except of a set of μ-measure zero, $F(x) > 0$ and $G(x) > 0$. Moreover, on the same set,

$$g_j(x) = w(x) f_j(x), \qquad j = 1, \dots, n, \qquad w(x) \ge 0\,, \tag{7.12}$$

where for $1 < p < +\infty$

$$w(x) = \begin{cases} F(x)^{p-2} & \text{if } \ F(x) \ne 0 \\ 0 & \text{if } \ F(x) = 0\,, \end{cases} \tag{7.13}$$

while for $p = \infty$,

$$\begin{cases} w(x) = 0 \ \ \text{outside of the set } \ A_0 = \{x : F(x) = 1 = \|F\|_\infty\}\,, \\ \displaystyle\int_A w \, d\mu = n\,. \end{cases} \tag{7.14}$$

Proof. For the maximal basis $\mathbf{f}$,

$$\begin{aligned} n = \Phi(\mathbf{f}) = \Psi(\mathbf{f}) &= \int_A \sum_{j=1}^{n} f_j g_j \, d\mu \le \int_A FG \, d\mu \\ &\le \|F\|_p \, \|G\|_{p'} = n\,. \end{aligned} \tag{7.15}$$

Both inequalities must be equalities. Our statements follow from exact descriptions of cases of equality in Hölder's inequalities. Thus, equality in the first inequality of (7.15) implies $g_j(x) = w(x)f_j(x)$, *a.e.*, $j = 1, \dots, n$, with some $w(x) \ge 0$. And the second inequality gives for $1 < p < \infty$, $G(x) = F(x)^{p-1}$ a.e.. Hence $wF = F^{p-1}$ and $w(x) = F(x)^{p-2}$ if $F(x) \ne 0$. If $F(x) = 0$, then all $f_j(x) = 0$, and we can replace $w(x)$ by zero.

If $p = \infty$, from the second equality, $G(x) = 0$ outside of A_0. We can assume that $w(x) = 0$ there. Moreover, since $\|F\|_\infty = 1$,

$$\int_A w\,d\mu = \int_{A_0} w\,d\mu = \int_{A_0} wF^2\,d\mu = \int_A FG\,d\mu = n\,. \qquad \square$$

Remark. For $1 < p < +\infty$, we can replace (7.12), (7.13) by the dual relations

$$(7.12a) \qquad f_j(x) = w_0(x)g_j(x)\,, \qquad w_0(x) \ge 0$$

$$(7.13a) \qquad w_0(x) = \begin{cases} G(x)^{p'-2} & \text{if} \quad G(x) \ne 0 \\ 0 & \text{if} \quad G(x) = 0\,. \end{cases}$$

We use Proposition 7.3, and for a while assume $2 \le p \le \infty$. For measurable functions on A, we define a Hilbert space H by means of the inner product

$$(7.16) \qquad (h_1, h_2)^* = \int_A wh_1h_2\,d\mu$$

and the norm

$$N(h) = \left(\int_A wh^2\,d\mu\right)^{1/2}.$$

Lemma 7.4. *We have the inclusion $L_p \subset H$ and the inequalities*

$$(7.17) \qquad N(h) \le n^{(1/2)-(1/p)}\|h\|_p\,, \qquad h \in L_p\,,$$

$$(7.18) \qquad \|h\|_p \le N(h)\,, \qquad h \in Y\,.$$

Proof. First we assume $2 \le p < +\infty$. Relation (7.17) follows from Hölder's inequality:

$$N(h)^2 = \int wh^2\,d\mu \le \left(\int (h^2)^{p/2}\,d\mu\right)^{2/p}\left(\int w^{p/(p-2)}\,d\mu\right)^{(p-2)/p}$$
$$= \|h\|_p^2\left(\int F^p\,d\mu\right)^{(p-2)/p} = n^{(p-2)/p}\|h\|_p^2\,.$$

On the subspace Y of L_p, the functions $f_1, \dots, f_n$ are an orthonormal basis with respect to the inner product (7.16):

$$\int_A wf_if_j\,d\mu = \int_A g_if_j\,d\mu = \phi_i(f_j) = \delta_{i,j}\,.$$

Using this, we have for $h \in Y$, $h = \sum_1^n a_if_i$ with $N(h)^2 = \sum a_i^2$. Then

$$\|h\|_p^p = \int |\sum a_if_i|^p\,d\mu \le \int |\sum a_if_i|^2\,S(x)^{p-2}\,d\mu\,,$$

where

$$S(x) := \sum_1^n |a_i|\,|f_i(x)| \le \left(\sum a_i^2\right)^{\frac{1}{2}}\left(\sum f_i(x)^2\right)^{\frac{1}{2}} = N(h)F(x)\,.$$

It follows that $S(x)^{p-2} \le N(h)^{p-2}w(x)$, and we obtain

$$\|h\|_p^p \le N(h)^{p-2}\int w\left(\sum a_if_i\right)^2 d\mu = N(h)^p\,.$$

In the case $p = +\infty$, we have

$$N(h)^2 = \int wh^2\,d\mu \le \|h\|_\infty^2\int w\,d\mu = n\|h\|_\infty^2\,,$$

$$\|h\|_\infty = \|\sum a_if_i\|_\infty \le \left(\sum a_i^2\right)^{\frac{1}{2}}\|F\|_\infty = N(h)\,,$$

and we have again the required inequalities. $\square$

We now arrive at one of our main theorems.

Theorem 7.5 (Lewis [1978]). *If Y is an n-dimensional subspace of $X = L_p$, $1 < p < \infty$, or of $X = C(A)$, $p = \infty$, then there is a projection of X onto Y with norm $\le n^{|(1/2)-(1/p)|}$.*

Proof. We show that the operator

$$(7.19) \qquad Ph := \sum_1^n (h, g_i)f_i\,, \qquad h \in X$$

(with f_i, g_i from Proposition 7.3) has the required properties. Obviously, P maps X onto Y and is a projection.

First let $2 \le p \le +\infty$. Let H be the Hilbert space introduced earlier. Then P is the restriction to X of the orthogonal projection of H onto Y, given by

$$(7.20) \qquad Ph := \sum_1^n (h, f_i)^*f_i\,, \qquad h \in H\,.$$

For $h \in L_p$, from (7.16) and (7.17),

$$\|Ph\|_p \le N(Ph) \le N(h) \le n^{(1/2)-(1/p)}\|h\|_p\,,$$

as desired.

The case $1 < p \le 2$ follows by duality by means of the remark to Proposition 7.3. Let $Y^* \subset L_{p'}$ be the space spanned by the g_i. The above argument shows that the operator $P^*k := \sum_1^n (k, f_i)' g_i$, $k \in L_{p'}$, projects $L_{p'}$ onto Y^* with norm $\le n^{(1/2)-(1/p')}$; here $(\cdot,\cdot)'$ is the inner product (7.16) with w replaced by some new function w_0. Since

$$\int_A kPh\, d\mu = \int_A hP^*k\, d\mu\,,$$

we get

$$\begin{aligned}\|Ph\|_p &= \sup_{\|k\|_{p'}\le 1} \int kPh\, d\mu \le \|h\|_p \sup_{\|k\|_{p'}\le 1} \|P^*k\| \\ &= \|h\|_p \|P^*\| \le n^{(1/2)-(1/p')}\|h\|_p = n^{|1/2-1/p|}\|h\|_p\,. \qquad \square\end{aligned}$$

Each Banach space can be isometrically imbedded into the space $C(A)$ for some compact Hausdorff space A. We obtain therefore a theorem of Kadec and Snobar [1971]:

Theorem 7.6. *For any n-dimensional subspace Y of a Banach space X, there exists a projection of X onto Y of norm $\le \sqrt{n}$.*

In particular, Theorem 7.6 is valid also for $p = 1$.

Theorem 5.5 shows that this cannot be improved, at least up to a constant factor. (See also Note **10.8**.) Likewise, Theorem 7.3 is the best possible. By means of matrix transformations, an n-dimensional subspace $Y \subset l_p^{2n}$, $1 < p < \infty$ can be constructed (see Sobczyk [1941]) for which each projection of l_p^{2n} onto Y has norm $\ge$ Const $n^{|1/2-1/p|}$.

§ 8. Families of Commuting Operators

For commuting operators V_n, $n = 1, 2, \dots$, that map a Banach space X onto itself, we shall develop approximation theorems that parallel those of §§2,5,6 of Chapter 7. However, instead of sequences, we shall concentrate our attention on the formally more general continuous families (U_t), $0 < t \le 1$, for $t \to 0+$. From a sequence (V_n) we obtain a family by putting $U_t = V_{[1/t]}$, $0 < t \le 1$.

The operators U_t commute if $U_tU_s = U_sU_t$ for all s, t, $0 < s, t \le 1$. A rich theory exists for *semi-groups of operators* U_t, $0 < t < \infty$. They have the property $U_tU_s = U_{t+s}$; this implies commutivity. See the books of Hille and Phillips [B-1957] and Butzer and Berens [A-1967].

Another example of commuting families is provided by convolution operators on $L_1(\mathbb{R})$ or $L_1(\mathbb{T})$. If $K \in L_1(\mathbb{R})$ is a given kernel, the operator

$$U(f,x) := (K * f)(x) = \int_{\mathbb{R}} K(x-t) f(t)\, dt \tag{8.1}$$

maps (by Fubini's theorem), the space $L_1(\mathbb{R})$ into itself. Now the space L_1 is a commutative (and associative) algebra under convolution, because for $f, g \in L_1$,

$$(f*g)(x) = \int_{\mathbb{R}} f(x-t)g(t)\, dt = \int_{\mathbb{R}} f(t)g(x-t)\, dt = (g*f)(x)\,.$$

Therefore, all operators of convolution type (8.1) commute with each other.

Theorems 8.1 to 8.3 here are due to Butzer and Scherer ([1972]). Predecessors are Berens [1964], for the case of semi-groups, and Sunouchi [1968], who proved a part of Theorem 8.1

Let X be a Banach space, Y one of its subspaces, dense in X, equipped with a semi-norm $|\cdot|_Y$. We assume that the commutative family (U_t) is uniformly bounded: $\|U_t\| \le M$, and that the U_t map X into Y. We measure the approximation of $f \in X$ by U_tf by means of a positive increasing function $\varphi(t)$, $0 < t \le 1$, $\varphi(t) \to 0$ for $t \to 0+$, and by a number $\gamma \ge 0$. The following are called the *Jackson* and the *Bernstein inequality of the family* (U_t) respectively:

(J) $\qquad \|U_tf - f\|_X \le Ct^\gamma |f|_Y\,, \qquad f \in Y\,,\ 0 < t \le 1\,,$

(B) $\qquad |U_tf|_Y \le Ct^{-\gamma}\|f\|_X\,, \qquad f \in X\,,\ 0 < t \le 1\,.$

We shall assume that φ and γ satisfy for $t \to 0+$ the following

$$\int_0^t u^{-1}\varphi(u)\, du = \mathcal{O}(\varphi(t))\,, \tag{8.2}$$

$$\int_t^1 u^{-1-\gamma}\varphi(u)\, du = \mathcal{O}(t^{-\gamma}\varphi(t))\,. \tag{8.3}$$

As a consequence of the last condition, $t^{-\gamma}\varphi(t)$ is bounded from below by a positive constant, for we have $t^{-\gamma}\varphi(t) \ge$ Const $\int_{1/2}^1 u^{-1-\gamma}\varphi\, du$ for $t \le \frac{1}{2}$. For example, $\varphi(u) = u^c$, $0 < c < \gamma$ satisfies these requirements, but u^γ does not. The following are easily derived from (8.2), (8.3) and the monotonicity of φ:

$$\sum_{k=0}^\infty \varphi(t2^{-k}) \le \varphi(t) + 2\int_0^t u^{-1}\varphi(u)\, du \le C\varphi(t)\,, \tag{8.4}$$

$$\sum_{1\le 2^k \le t^{-1}} 2^{k\gamma}\varphi(2^{-k}) \le \varphi(1) + 2^{1+\gamma}\int_t^1 u^{-1-\gamma}\varphi(u)\, du \le Ct^{-\gamma}\varphi(t)\,. \tag{8.5}$$

Theorem 8.1 (Butzer and Scherer). *Let U_t be a uniformly bounded family of commuting operators that map X into Y and satisfy* (J) *and* (B). *Then for any element $f \in X$ the following statements are equivalent as $t \to 0+$:*

(i) $\|U_tf - f\|_X = \mathcal{O}(\varphi(t))$,

(ii) $K(f, t^\gamma; X, Y) = \mathcal{O}(\varphi(t))$,

(iii) $|U_tf|_Y = \mathcal{O}(t^{-\gamma}\varphi(t))$.

Proof. (a) (iii) *implies* (i). For $0 < u, v \le 1$ we have, using (J),

$$\begin{aligned}\|U_uf - U_vf\|_X &\le \|(1-U_v)U_uf\|_X + \|(1-U_u)U_vf\|_X \\ &\le C(v^\gamma|U_uf|_Y + u^\gamma|U_vf|_Y)\,.\end{aligned} \tag{8.6}$$

In particular, for $u = t2^{-k}$, $v = t2^{-k-1}$, by (iii)

$$(8.7)\qquad \begin{aligned} \|U_{t2^{-k}}f - U_{t2^{-k-1}}f\|_X &\le Ct^\gamma 2^{-k\gamma}(2^{-\gamma}|U_{t2^{-k}}f|_Y + |U_{t2^{-k-1}}f|_Y) \\ &\le C_1 t^\gamma 2^{-k\gamma}(t2^{-k})^{-\gamma}\varphi(t2^{-k}) = C_1\varphi(t2^{-k}). \end{aligned}$$

Therefore, the series

$$U_t f - f = \sum_{k=0}^{\infty}(U_{t2^{-k}}f - U_{t2^{-k-1}}f)$$

converges, and from (8.7) and (8.4) we obtain $\|U_t f - f\|_X \le \text{Const}\ \varphi(t)$.

(b) (i) *implies* (iii). Similarly to (8.6), using first (B) and then (i) we derive, for $0 < v \le u \le 1$,

$$(8.8)\qquad \begin{aligned} |U_u f - U_v f|_Y &\le |U_u(I - U_v)f|_Y + |U_v(I - U_u)f|_Y \\ &\le Cu^{-\gamma}\|U_v f - f\|_X + Cv^{-\gamma}\|U_u f - f\|_X \\ &\le C_1\big(u^{-\gamma}\varphi(v) + v^{-\gamma}\varphi(u)\big) \le 2C_1 v^{-\gamma}\varphi(u). \end{aligned}$$

For an arbitrary $t \in (0,1]$ we choose an integer n so that $2^{-n-1} < t \le 2^{-n}$. Then

$$(8.9)\qquad U_t f = (U_t f - U_{2^{-n-1}}f) + U_1 f - \sum_{k=0}^{n}(U_{2^{-k}}f - U_{2^{-k-1}}f).$$

In view of (8.8), (8.5) and (8.3), since $|U_1 f|_Y \le CM$,

$$\begin{aligned} |U_t f|_Y &\le 2C_1\Big\{2^{(n+1)\gamma}\varphi(t) + |U_1 f|_Y + \sum_{k=0}^{n} 2^{(k+1)\gamma}\varphi(2^{-k})\Big\} \\ &\le C_2\Big\{t^{-\gamma}\varphi(t) + |U_1 f|_Y + \varphi(1) + 2^{1+\gamma}\int_t^1 u^{-1-\gamma}\varphi(u)\,du\Big\} \\ &= \mathcal{O}\big(t^{-\gamma}\varphi(t)\big). \end{aligned}$$

(c) *We show that* (i) *and* (iii) *together are equivalent to* (ii). For a given $f \in X$ and for arbitrary $g \in Y$ we find, using the uniform boundedness of the norms and (J):

$$\begin{aligned} \|U_t f - f\|_X &\le \|(U_t - I)(f - g)\|_X + \|(U_t - I)g\|_X \\ &\le (M+1)\|f - g\|_X + Ct^\gamma|g|_Y \\ &\le C_1\big(\|f - g\|_X + t^\gamma|g|_Y\big). \end{aligned}$$

We take the infimum for all g and obtain, by the definition of the K-functional,

$$\|U_t f - f\|_X \le C_1 K(f, t^\gamma; X, Y),$$

so that (ii) implies (i).

On the other hand, taking $g = U_t f$,

$$K(f, t^\gamma; X, Y) \le \|f - U_t f\|_X + t^\gamma|U_t f|_Y,$$

and if f satisfies both (i) and (iii), we derive (ii). □

There is a similarity between the assumptions and statements of Theorem 8.1 and those of Peetre's Theorem 5.1. This can be made more formal. For a sequence of uniformly bounded operators $(U_n)_1^\infty$ on a Banach space X we define the maximal function of the U_n by

$$(8.10)\qquad M_n(f) := \sup_{m \ge n}\|f - U_m f\|_X, \quad f \in X, \ n = 1, 2, \dots.$$

Theorem 8.2. (i) *If the U_n satisfy* (J), *then for $K(f,t) := K(f,t;X,Y)$*

$$(8.11)\qquad \|f - U_n f\| \le CK(f, n^{-\gamma}), \qquad n = 1, 2, \dots.$$

(ii) *If the U_n commute and satisfy* (B), *then*

$$(8.12)\qquad K(f, n^{-\gamma}) \le Cn^{-\gamma}\Big\{\|f\|_X + \sum_{k=1}^{n} k^\gamma M_k(f)\frac{1}{k}\Big\}.$$

The first part has been shown in (c) above. The second part can be easily derived following the argument of (8.9). Of course, Theorem 8.2 can be used to replace parts of the proof of Theorem 8.1. We leave the details to the reader.

The main point of Theorem 8.1 is as follows. From the direct theorem (J), with error of order t^γ, we derive a family of intermediate theorems with larger errors $\mathcal{O}(\varphi(t))$. One gets characterizations of all spaces $\{f : \|f - U_t f\| \le C\varphi(t)\}$ with the specified φ. Besides the commutivity of the U_t, one needs for this purpose the condition (B), which is sometimes easy to prove. Thus, if $U_n f \in \mathcal{T}_n$ is a trigonometric polynomial of degree $\le n$, and $\gamma = r = 1, 2, \dots$, $X = L_p(\mathbb{T})$, $Y = W_p^r(\mathbb{T})$, then (B) follows for the Bernstein inequality for $\mathcal{T}_n$ and the uniform boundednes of the U_n.

We do not claim that Theorem 8.1 is valid for $\varphi(t) = t^\gamma$. For this function φ, (i) may well be a saturation condition. The proof of (ii) then usually requires different arguments. Saturation spaces will be discussed in Chapter 11. Intermediate theorems for Bernstein polynomials can be found in §5 and §8 of Chapter 10.

Remark. Examining part (a) of the proof of Theorem 8.1 we see that we have established (i) for some $f \in X$ under the following conditions: f satisfies (iii); the Jackson inequality (J) is valid for all $f \in X$ with the exponent $\gamma > 0$, while the function φ satisfies (8.2). Moreover, the monotonicity of φ has been used moderately in the proof: the argument works for any ψ instead of φ for which

$$(8.13)\qquad \psi(t) \le C\psi(2t), \qquad 0 < t \le \tfrac{1}{2}.$$

An example of the spaces X, Y of the theorem is provided by $X = L_p$, $Y = W_p^r$, $1 \le p \le \infty$, $r = 1, 2, \dots$. In this case, a natural space between X and Y is $Z := W_p^k$, $1 \le k < r$. The question arises whether we can prove (i) for $f \in Z$ and an appropriate error. This can indeed be done.

Let $Y \subset Z \subset X$, where Z is a linear subspace of X equipped with a semi-norm $|\cdot|_Z$ and the norm $\|\cdot\|_Z = \|\cdot\|_X + |\cdot|_Z$. We shall assume that Z satisfies the Bernstein inequality for (U_t), the spaces Z and X, and the exponent δ, $0 < \delta < \gamma$:

$$(\mathrm{B})_\delta \qquad |U_t f|_Z \le C t^{-\delta} \|f\|_X \,.$$

In addition we assume that

$$(8.14) \qquad \int_0^t u^{-1-\delta} \varphi(u)\, du = \mathcal{O}\big(t^{-\delta}\varphi(t)\big) \,.$$

(It is easy to see that this implies (8.2).) For simplicity we shall assume that Z is complete.

We shall prove a Jackson inequality for the spaces Y, Z:

Proposition 8.3. *If the spaces Y, X satisfy the Jackson inequality* (J), *and the spaces Z, X satisfy the Bernstein inequality* $(\mathrm{B})_\delta$, *then Y, Z satisfy the Jackson inequality with exponent $\gamma-\delta$:*

$$(\mathrm{J})_{\gamma-\delta} \qquad \|U_t f - f\|_Z \le C t^{\gamma-\delta} |f|_Y \,, \quad \textit{for} \quad f \in Y \,.$$

Proof. Let $f \in Y$. Arguing as in (8.8) for the semi-norm $|\cdot|_Z$ instead of $|\cdot|_Y$, with the help of $(\mathrm{B})_\delta$, we obtain

$$|U_t f - f|_Z \le \sum_{k=0}^{\infty} |U_{t2^{-k}} f - U_{t2^{-k-1}} f|_Z \le C \sum_{k=0}^{\infty} (t2^{-k})^{-\delta} \|U_{t2^{-k}} f - f\|_X \,.$$

By means of (J) we see that this is

$$\le C \sum_{k=0}^{\infty} (t2^{-k})^{-\delta} (t2^{-k})^{\gamma} |f|_Y = \mathrm{Const}\ t^{\gamma-\delta} |f|_Y \,.$$

This, together with (J) and the definition of the norm $\|\cdot\|_Z$, yields $(\mathrm{J})_{\gamma-\delta}$. □

Theorem 8.4. *Let (U_t) be a uniformly bounded family of commuting operators from X to Y which satisfies* (J), (B) *and* $(\mathrm{B})_\delta$ *with parameters $0 < \delta < \gamma$ and with a function φ that has the properties* (8.3) *and* (8.14). *Then for each element $f \in X$ that satisfies one of the conditions* (i)–(iii), *one has also*

$$(\mathrm{iv}) \qquad \|U_t f - f\|_Z = \mathcal{O}\big(t^{-\delta}\varphi(t)\big) \,.$$

Proof. We assume that $f \in X$ satisfies (iii), which we rewrite in the form

$$(8.15) \qquad |U_t f|_Y = \mathcal{O}\big(t^{-\gamma+\delta}\psi(t)\big) \,,$$

where $\psi(t) := t^{-\delta}\varphi(t)$ satisfies (8.13). By (8.14), we also have $\int_0^t \frac{1}{u}\psi(u)\, du = \mathcal{O}(\psi(t))$. Together with $(\mathrm{J})_{\gamma-\delta}$, we have all requirements of the Remark after Theorem 8.1 for the spaces Y, Z and the function ψ. Therefore we have (i) for these spaces, that is, $\|U_t f - f\|_Z = \mathcal{O}(\psi(t))$, which is (iv). □

As a simple illustration of our theorems, we consider the Jackson integrals $J_n(f)$ of §2 of Chapter 7. We take $X = L_p(\mathbb{T})$, $Z = W_p^1(\mathbb{T})$, $Y = W_p^2(\mathbb{T})$, $\gamma = 2$, $\varphi(t) = t^\alpha$. The condition (J): $\|f - J_n f\|_p \le C n^{-2} \|f\|_p$ follows from Theorem 2.2 of Chapter 7, also (B) is satisfied. Proposition 8.2 yields $\|(J_n f)' - f'\|_p \le C n^{-1} \|f''\|_p$, and we get:

Theorem 8.5. *For the Jackson operator $J_n(f)$ the following statements are equivalent if $0 < \alpha < 2$ for $f \in L_p(\mathbb{T})$:*

$$(8.16) \qquad \begin{cases} \|f - J_n f\|_p = \mathcal{O}(n^{-\alpha}) \\ \omega_2(f, t)_p = \mathcal{O}(t^\alpha) \\ \|(J_n f)''\|_p = \mathcal{O}(n^{2-\alpha}) \end{cases}$$

If $1 < \alpha < 2$, then each of the conditions (8.16) *is equivalent to*

$$(8.17) \qquad \|f - (J_n f)'\|_p = \mathcal{O}(n^{-\alpha+1}) \,.$$

More difficult to handle is the *de la Valleé-Poussin integral*

$$(8.18) \qquad \begin{cases} V_n(f, x) := \dfrac{1}{\pi} \displaystyle\int_{\mathbb{T}} f(x-t) K_n(t)\, dt \,, \\ K_n(t) := \dfrac{1}{2} \dfrac{n!}{(2n)!} \Big(2\cos\dfrac{t}{2}\Big)^{2n} . \end{cases}$$

The kernel $K_n(t)$ is a positive, even trigonometric polynomial. With $2\cos\frac{t}{2} = e^{it/2} + e^{-it/2}$ the binomial formula yields

$$(8.19) \qquad K_n(t) = \frac{1}{2} + \sum_{k=1}^{n} \rho_n(k) \cos kt \,, \qquad \rho_n(k) = \frac{(n!)^2}{(n-k)!(n+k)!} \,.$$

By means of the arguments that have been used to prove (2.7) of Chapter 7, or by means of §2 of Chapter 11, one establishes that for $f \in L_p$, $1 \le p \le \infty$,

$$(8.20) \qquad \|V_n f - f\|_p \le C \omega_2\Big(f, \frac{1}{\sqrt{n}}\Big)_p \,.$$

Another important estimate is

$$(8.21) \qquad \|(V_n f)''\|_p \le n \|f\|_p \,, \qquad f \in L_p \,.$$

For the proof, we write the integral (8.18) in the form

$$(8.22)\qquad \begin{cases} V_n(f,x) = \dfrac{1}{2\pi}\displaystyle\int_{\mathbb{T}} f(u)v_n(x-u)\,du \\ v_n(x) = \displaystyle\sum_{k=-\infty}^{+\infty} \widehat{v}_n(k)e^{ikx} \end{cases}$$

where $\widehat{v}_n(k) = \rho_n(k)$ (see (8.19)), for $|k| \le n$, $= 0$ for $|k| > n$. Then

$$v_n''(x) = -\sum_{k=-\infty}^{\infty} k^2\widehat{v}_n(k)e^{ikx}.$$

We need the identities

$$(8.23)\qquad \begin{cases} k^2\widehat{v}_n(k) = n^2\big(\widehat{v}_n(k) - \widehat{v}_{n-1}(k)\big), & k,n\in\mathbb{Z} \\ n^2\big(v_{n-1}(x) - v_n(x)\big) = \dfrac{n^2}{2}v_{n-1}(x)(1-\cos x) - \dfrac{n}{2}v_n(x). \end{cases}$$

From this, differentiating (8.22) we get

$$V_n(f,x)'' = \frac{1}{2\pi}\int_{\mathbb{T}} f(x-u)v_n''(u)\,du = (f * g_n)(x),$$

where $g_n(x) := -\frac{n^2}{2}v_{n-1}(x)(1-\cos x) + \frac{n}{2}v_n(x)$. Since the v_n are positive, (8.23) yields

$$\begin{aligned}\|(V_nf)''\|_p &\le \frac{1}{2\pi}\|g_n\|_1\|f\|_p \\ &\le \left(\frac{n^2}{2}\|v_{n-1}(x)(1-\cos x)\|_1 + \frac{n}{2}\|v_n\|_1\right)\|f\|_p \\ &= \left(\frac{n^2}{2}(1-\widehat{v}_{n-1}(1)) + \frac{n}{2}\right)\|f\|_p = n\|f\|_p,\end{aligned}$$

proving (8.21).

We take $X = L_p(\mathbb{T})$, $Y = W_p^2(\mathbb{T})$, $\gamma = 1$, $\varphi(t) = t^\alpha$, $0 < \alpha \le 1$. Then (8.20) implies Jackson's inequality (J), and (8.21) is the Bernstein inequality (B). We obtain

Theorem 8.6. *For the de la Valleé-Poussin operators* (8.16), *the following statements are equivalent for* $0 < \alpha < 1$*:*

$$(8.24)\qquad \begin{cases} \|V_nf - f\|_p = \mathcal{O}(n^{-\alpha}) \\ \omega_2(f,t) = \mathcal{O}(t^{2\alpha}) \\ \|(V_nf)''\|_p = \mathcal{O}(n^{1-\alpha}). \end{cases}$$

In Chapter 11, we shall see that in both of our examples, (J) is the saturation condition.

§ 9. Problems

9.1. Prove a simpler version of Theorem 2.3, namely $\Lambda_n \le 2\log n + 4$ by observing that $|l_{n,k}(x)|$ in (2.9) does not exceed 2 and also does not exceed $\frac{1}{\pi|t-t_k|}$.

9.2. Prove that the Lebesgue constant for the trigonometric interpolation at $2n+1$ equidistant knots on $\mathbb{T}$ is equal to

$$\frac{1}{2n+1}\left\{1 + 2\sum_{k=0}^{n-1}\left(\sin\frac{(2k+1)\pi}{2(2n+1)}\right)^{-1}\right\}$$

(Ehlich und Zeller [1966]).

9.3. State and prove the theorem that corresponds to Theorem 5.1 for the test functions $\phi_0 = 1$, $\phi_1 = \cos t$, $\phi_2 = \sin t$ on $\mathbb{T}$.

9.4. In the situation of Theorem 4.2 prove that if $f \in C^1[-1,+1]$, then

$$\|f - U_nf\| \le \mathrm{Const}\left(\|f\|\lambda_n + \|f'\|\lambda_n + \sqrt{\lambda_n}\,\omega(f', \sqrt{\lambda_n})\right).$$

Use the relations $f(x) = f(a) + f'(a)(x-a) + R(x)$, $|R(x)| \le |x-a|\omega(f', |x-a|)$, but not K-functionals.

9.5. If Y is an n-dimensional subspace of a Banach space X, $\mathbf{f} = (f_1,\dots,f_n)$ a basis for Y, subject to $\|f_j\| \le 1$, $j = 1,\dots,n$, and $\varphi = (\varphi,\dots,\varphi_n)$ a fixed basis for Y^*, then there exists an $\mathbf{f}$ for which $\det[\varphi_i(f_j)]_{i,j=1}^n$ achieves its maximum. For the extremal $\mathbf{f}$ and the dual basis $\psi = (\psi_1,\dots,\psi_n)$ one has $\|f_i\| = \|\psi_i\| = 1$, $i = 1,\dots,n$ (Auerbach).

9.6. Using Problem 9.5, prove the existence of a projection P of X onto Y with $\|P\| \le n$.

9.7. If a maximal basis $\mathbf{f}$ for $Y \subset L_p$ is known, describe all other maximal bases $\mathbf{F}$ of Y. [Answer: They are given by $\mathbf{F} = T\mathbf{f}$, where T is any non-singular $n \times n$ matrix for which $\det T = \det T^{-1} = 1$.]

§ 10. Notes

10.1. An old conjecture by Bernstein and Erdös was that the interpolation points x_k, $k = 1,\dots,n$, in $[-1,1]$ which produce the smallest possible norm Λ_n^* of the Lagrange interpolation formula (2.6) are unique and that their Lebesgue function satisfies $\lambda_n(x_k) = \Lambda_n^*$, $k = 1,\dots,n$. This has been proved only recently, see Kilgore [1978], also deBoor and Pinkus [1978].

10.2. There are only a few books dealing with interpolation theory: Natanson [A-1954, vol. 3]. The following books deal with special aspects of the theory: Lorentz, Jetter and Riemenschneider [A-1983], Szabados and Vértesi [A-1990], R.A. Lorentz [A-1992].

10.3. Estimates similar to (4.6) exist also for the L_p-spaces; one takes here $\lambda_n = \max_{i=0,1}\|\phi_i - U_n\phi_i\|$ without the use of the function ϕ_2. See Berens and DeVore [1976], [1978].

10.4. One can show that the Fourier projection in Theorem 6.1 is the *unique* projection with the smallest norm. Unfortunately, the proof is not simple (Cheney, Hobby, Morris, Schurer and Wulbert [1969]). Fischer, Morris and Wulbert [1981] have further theorems of this type. Let $A(D)$, $D = \{z : |z| \le 1\}$ be the space of functions

$f(z) = \sum_{k=0}^{\infty} a_k z^k$, analytic for $|z| < 1$ and continuous on D, equipped by the uniform norm. As early as 1915, Landau (see Landau-Gaier [B-1986, p. 26]) gave the exact values of the norms of the projections of $A(D)$ onto $\mathcal{P}_n$, given by the Taylor sums $S_n(f,z) := \sum_{k=0}^{n} a_k z^k$:

$$\|S_n\| = \left(\frac{1}{2}\right)^2 + \cdots + \left(\frac{1.3\ldots(2n-1)}{2.4\ldots 2n}\right)^2 \approx \frac{1}{\pi}\log n\,.$$

10.5. Let us assume that for some continuous function f, the error of approximation $E_n(f)$ is exactly of order $n^{-\alpha}$, $\alpha > 0$. One would expect that a modest amount of regularity of f would imply the existence of the limit $\mu = \lim_{n\to\infty} n^\alpha E_n(f)$. But this is known only in a few cases. In Bernstein [A-1952, vol. 1, p. 157, vol. 2, p. 262] this is proved for $f(x) = |x|^p$, $p > 0$, $x \in [-1,1]$. Bernstein's computations are based upon the relation $n^p E_n(|x|^p) - E_n(f_p(nx)) \to 0$, $n \to \infty$, where

$$f_p(t) = -\frac{4}{p}\sin\frac{\pi p}{2}\ \text{Const}\int_0^\infty \frac{u^{p+1}\,du}{(e^u - e^{-u})(u^2+t^2)}\,.$$

10.6. Integrating by parts the formula (10.14) of Chapter 3, Fiedler and Jurkat [1983] obtain *asymptotic expansion* for $E_n(f)_1$ in cases when f is a series $f(z) = \sum a_n z^n$ with a radius of convergence > 1, and $f^{(n)}(x) \ge 0$, $-1 \le x \le 1$, $n = 0, 1, \ldots$. Then

$$E_n(f)_1 = \pm \int_{-1}^{+1} f^{(n)}(t) V_n(t)\,dt\,,$$

where $V_n(t)$ has a simple expression, related to the Chebyshev polynomial of the second kind. Therefore,

$$E_n(f)_1 = \pm \sum_{k=0}^{\infty} c_{n,k} a_{n+2k}\,,$$

and one can get good asumptotic formulas for the coefficients $c_{n,k}$.

10.7. Many older papers on Fourier series deal with operators given by summability methods. We give only one example (by Nikolskii, see Timan [A-1963, p. 504]). If the $\lambda_k^{(n)}$, $k = 0, \ldots, n$, are convex or are concave for each n, then

$$\frac{a_0}{2}\lambda_0^{(n)} + \sum_{k=0}^{n} \lambda_k^{(n)}(a_k \cos kx + b_k \sin kx) \to f(x)\,, \qquad x \in \mathbb{T}\,,$$

for all continuous functions f if and only if

(i) $\lambda_k^{(n)} \to 1$ for $n \to \infty$, and each $k = 0, 1, \ldots$

(ii) $\lambda_k^{(n)} = \mathcal{O}(1)$, $k, n = 0, 1, 2, \ldots$

(iii) $\sum_0^n \frac{1}{n-k+1}\lambda_k^{(n)} = \mathcal{O}(1)$, $n = 1, 2, \ldots$.

10.8. For an n-dimensional Banach space X, let $\lambda_n(X)$ be the infimum of norms of the projections onto X of Banach spaces $Y \supset X$, into which X is continuously imbedded. By the theorem of Kadec and Snobar, $\lambda_n(X) < \sqrt{n}$ for each X. König and Tomczak-Jaegermann [1990] prove more

$$\lambda_n(X) = \sqrt{n} - \frac{1}{2\sqrt{n}} + \mathcal{O}(n^{-3/4})\,.$$

Chapter 10. Bernstein Polynomials

§ 1. Definitions and Inequalities

The simplest properties of the Bernstein polynomials

$$(1.1)\qquad B_n(f,x) := \sum_{\nu=0}^{n} f\left(\frac{\nu}{n}\right) p_{n,\nu}(x)\,, \qquad p_{n,\nu}(x) := \binom{n}{\nu} x^\nu (1-x)^{n-\nu}$$

have been discussed in Chapter 1. The book of Lorentz [A-1986] is devoted to results on Bernstein polynomials (before 1951) for functions on $[0,1]$ and $\mathbb{C}$, together with applications. The main content of this chapter are later results, about approximation and shape preserving properties of the B_n, and the saturation and inverse theorems.

Let $A := [0,1]$ throughout this chapter. The $B_n(f)$ are linear positive operators which reproduce linear functions l: $B_n(l) = l$, and are defined for each bounded function f on A. We shall assume in most cases that $f \in C(A)$. If one replaces $f(\nu/n)$ in (1.1) by some average $f_{n,\nu}$ of f on an interval close to ν/n, one obtains similar operators. The choice

$$f_{n,\nu} = (n+1)\int_{\frac{\nu}{n+1}}^{\frac{\nu+1}{n+1}} f(t)\,dt\,,$$

gives the *Kantorovich polynomials* $K_n(f,x)$, defined for each $f \in L_1(A)$.

Sums of the $p_{n,\nu}$ can be estimated, in an elementary way, by means of the expressions

$$(1.2)\qquad T_{n,s} := \sum_{\nu=0}^{n} (\nu - nx)^s p_{n,\nu}(x)\,, \qquad n = 1, 2, \ldots\,,\ s = 0, 1, \ldots\,.$$

From (1.7) of Chapter 1, we have $T_{n,0}(x) = 1$, $T_{n,1}(x) = 0$. Further information is given by the following theorem.

Theorem 1.1. *For a fixed $s = 0, 1, \ldots$, $T_{n,s}(x)$ is a polynomial in x of degree $\le s$, and in n of degree $[s/2]$. Moreover, if $X := x(1-x)$, then*

$$(1.3)\qquad T_{n,2s}(x) = \sum_{j=1}^{s} a_{j,s}(X) n^j X^j\,,$$

$$T_{n,2s+1}(x) = (1-2x)\sum_{j=1}^{s} b_{j,s}(X)n^j X^j \tag{1.4}$$

where $a_{j,s}$, $b_{j,s}$ are polynomials of degree $\le s-j$, with coefficients independent of n.

Proof. We have

$$\begin{aligned} T'_{n,s}(x) &= -nsT_{n,s-1}(x) + \sum_{\nu=0}^{n}(\nu - nx)^s p'_{n,\nu}(x) \\ &= -nsT_{n,s-1}(x) + \sum_{\nu=0}^{n}(\nu-nx)^{s+1}\binom{n}{\nu}x^{\nu-1}(1-x)^{n-\nu-1}. \end{aligned}$$

From this we deduce the recurrence formula

$$T_{n,s+1}(x) = x(1-x)\left[T'_{n,s}(x) + nsT_{n,s-1}(x)\right].$$

All the statements of Theorem 1 now follow easily by induction from this. □

Some examples of $T_{n,s}$ are

$$T_{n,2} = nX, \qquad T_{n,3} = n(1-2x)X, \qquad T_{n,4} = 3n^2X^2 + n(X - 6X^2).$$

From Theorem 1.1, we can derive some useful estimates. For each $s = 0, 1, \dots,$ there is a constant A_s such that

$$0 \le T_{n,2s}(x) \le A_s n^s. \tag{1.5}$$

A corollary is that for each $\delta > 0$, and $s = 1, 2, \dots,$ there is a constant $C = C(\delta, s)$ for which

$$\sum_{|\nu/n - x|\ge\delta} p_{n,\nu}(x) \le Cn^{-s}, \qquad n = 1, 2, \dots. \tag{1.6}$$

Indeed, the sum does not exceed

$$\delta^{-2s}\sum_{|\nu/n-s|\ge\delta}\left(\frac{\nu}{n} - x\right)^{2s} p_{n,\nu}(x) \le \delta^{-2s}n^{-2s}T_{n,2s}(x) \le A_s\delta^{-2s}n^{-s}.$$

Another remark is that if $h(y) \ge 0$ is bounded for $0 \le y \le 1$ and converges to zero with y, then for any $r = 0, 1, \dots,$ uniformly in x,

$$n^r\sum_{\nu=0}^{n} h\left(\left|\frac{\nu}{n} - x\right|\right)\left(\frac{\nu}{n} - x\right)^{2r} p_{n,\nu}(x) \to 0. \tag{1.7}$$

If $\varepsilon > 0$, we can take $\delta > 0$ so that $|h(y)| < \varepsilon$ for $0 \le y \le \delta$. The portion of the sum (1.7) for $|\nu/n - x| < \delta$ is $\le \varepsilon n^{-r}T_{n,2r}(x) \le A_r\varepsilon$ and the portion for $|\nu/n - x| \ge \delta$ converges to zero by (1.6) with $s = r + 1$.

We can evaluate the integrals of the $p_{n,\nu}$ with the help of the Euler Γ function:

$$\begin{aligned} \int_0^1 p_{n,\nu}(x)\,dx &= \binom{n}{\nu}\int_0^1 x^\nu(1-x)^{n-\nu}\,dx = \binom{n}{\nu}\frac{\Gamma(\nu+1)\Gamma(n-\nu+1)}{\Gamma(n+2)} \\ &= \binom{n}{\nu}\frac{\nu!(n-\nu)!}{(n+1)!} = \frac{1}{n+1}. \end{aligned} \tag{1.8}$$

The next identity will be useful in §6. We define $S_k := \sum_{j=1}^k 1/j$, $k = 1, 2, \dots,$ $S_0 := 0$. Then,

$$\sum_{k=0}^{n}(S_n - S_k)p_{n,k}(x) = \sum_{k=1}^{n}\frac{(1-x)^k}{k}, \qquad n = 0, 1, \dots. \tag{1.9}$$

To prove (1.9), from

$$S_k = \sum_{j=0}^{k-1}\int_0^1 u^j\,du = \int_0^1 \frac{1-u^k}{1-u}\,du,$$

we derive that $S_n - S_k = \int_0^1 \frac{u^k - u^n}{1-u}\,du$ and obtain

$$\begin{aligned} \sum_{k=0}^{n}(S_n - S_k)p_{n,k}(x) &= \int_0^1 \frac{((1-x)+xu)^n - u^n}{1-u}\,du \\ &= \int_0^1 \frac{((1-x)(1-u)+u)^n - u^n}{1-u}\,du \\ &= \sum_{k=1}^{n}\binom{n}{k}(1-x)^k\int_0^1 u^{n-k}(1-u)^{k-1}\,du. \end{aligned}$$

Then, we apply (1.8).

§ 2. Derivatives of Bernstein Polynomials

For the derivatives of the polynomials $p_{n,\nu}$ we can use the two formulas

$$p'_{n,\nu}(x) = n[p_{n-1,\nu-1}(x) - p_{n-1,\nu}(x)] = \frac{1}{x(1-x)}(\nu - nx)p_{n,\nu}. \tag{2.1}$$

For the purpose of this formula, $p_{n,-1} := p_{n,n+1} := 0$ for all n. We shall continue to write $X = x(1-x)$, and further use $\Delta f(x) := f(x + 1/n) - f(x)$ for the first difference of f with step $1/n$. The k-th iterate of Δ is denoted by Δ^k. Then the differentiation of (1.1) yields for $B'_n(f) := (B_n(f))'$:

$$B'_n(f, x) = n\sum_{\nu=0}^{n-1}\Delta f\left(\frac{\nu}{n}\right)p_{n-1,\nu} = \frac{1}{X}\sum_{\nu=0}^{n} f\left(\frac{\nu}{n}\right)(\nu - nx)p_{n,\nu}(x). \tag{2.2}$$

Iterating the first expression, we have

$$B_n^{(k)}(f,x) = n(n-1)\dots(n-k+1)\sum_{\nu=0}^{n-k} \Delta^k f\left(\frac{\nu}{n}\right) p_{n-k,\nu}(x)\,. \tag{2.3}$$

In particular,

$$B_n^{(k)}(f,0) = n(n-1)\dots(n-k+1)\Delta^k f(0)\,, \qquad k=0,\dots,n\,.$$

This gives the Taylor expansion:

$$B_n(f,x) = \sum_{k=0}^{n} \binom{n}{k} \Delta^k f(0) x^k\,. \tag{2.4}$$

If f is a polynomial of degree m, then $\Delta^k f(0) = 0$ for $k > m$ and $\Delta^m f(0) \neq 0$. Therefore, *the Bernstein polynomial of degree m is itself a polynomial of degree* m.

From the established formulas, we can derive the following convergence results (see Lorentz [1937], also a chapter on Bernstein polynomials by Lorentz in the book of Besikovich [A-1939] where the expressions $T_{n,s}$ were introduced).

Theorem 2.1. (i) *If $f \in C^k(A)$, then $\|B_n^{(k)}(f) - f^{(k)}\| \to 0$; (ii) if $f \in C(A)$, $0 \le x_0 \le 1$, and f is $(k-1)$ times continuously differentiable in some neighborhood of x_0, then $B_n^{(k)}(f,x_0) \to f^{(k)}(x_0)$ whenever $f^{(k)}(x_0)$ exists.*

Proof. (i) It is sufficient to note that for some ξ_ν with $\nu/n < \xi_\nu < (\nu+k)/n$,

$$\begin{aligned} n\dots(n-k+1)\Delta^k f\left(\frac{\nu}{n}\right) &= (1+o(1))n^k \Delta^k f\left(\frac{\nu}{n}\right) \\ &= (1+o(1)) f^{(k)}(\xi_\nu) = (1+o(1)) f^{(k)}(\nu/n) \end{aligned}$$

uniformly. Therefore, from (2.3), $B_n^{(k)}(f,x) = B_{n-k}(f^{(k)},x) + o(1) \to f^{(k)}(x)$ uniformly on A.

(ii) If $x_0 = 0$, $B_n^{(k)}(f,0) = n\cdots(n-k)\Delta^k f(0) \to f^{(k)}(0)$. Similarly for $x_0 = 1$. If $0 < x_0 < 1$, we use the following fact (see Proposition 4.4 of Chapter 4). For $i,j,k = 0,1,\dots$, and $2i+j \le k$, there exist polynomials $q_{i,j,k}$, which do not depend on ν and n and satisfy

$$\begin{aligned} &\frac{d^k}{dx^k}[x^\nu(1-x)^{n-\nu}] = Q(x)x^{\nu-k}(1-x)^{n-\nu-k}\,, \\ &Q(x) = \sum_{2i+j\le k} n^i(\nu-nx)^j q_{i,j,k}(x)\,. \end{aligned} \tag{2.5}$$

To prove (ii), we write $f = P + g$, where P is the Taylor polynomial to f of degree k at x_0 and g is of the form $g(x) = \varepsilon(x)(x-x_0)^k$ with $\varepsilon(x) \to 0$ for $x \to x_0$. For P, we may use (i). Since $B_n^{(k)}(f) = B_n^{(k)}(P) + B_n^{(k)}(g)$, it

is sufficient to show that the last derivative converges to zero for $n \to \infty$ at $x = x_0$. By the lemma, we have

$$B_n^{(k)}(g,x) = X^{-k} \sum_{2i+j\le k} q_{i,j,k}(x) \sum_{\nu=0}^{n} \varepsilon\left(\frac{\nu}{n}\right)\left(\frac{\nu}{n}-x_0\right)^k n^i(\nu-nx)^j p_{n,\nu}(x)\,. \tag{2.6}$$

Since $k+j \ge 2(i+j)$, the absolute value of the inner sum for $x = x_0$ does not exceed

$$n^{i+j}\sum_{\nu=0}^{n}\left|\varepsilon\left(\frac{\nu}{n}\right)\right|\left(\frac{\nu}{n}-x_0\right)^{2(i+j)} p_{n,\nu}(x_0)$$

and this tends to 0 because of (1.7). □

From the first of the representations (2.2), we obtain the relation between the Kantorovich and the Bernstein polynomials:

$$\frac{d}{dx} B_{n+1}(F,x) = K_n(f,x) \tag{2.7}$$

where F is an indefinite integral of $f \in L_1(A)$. Then, (ii) of Theorem 2.1 yields:

(2.8) *one has $K_n(f,x) \to f(x)$ at any point x where $f(x)$ is the derivative of its indefinite integral, in particular, almost everywhere.*

§ 3. Approximation and Shape Preserving Properties

We begin with the following simple but important fact (Voronovskaja [1932]).

Theorem 3.1. *If f is bounded on A, differentiable in some neighborhood of x, and has second derivative $f''(x)$ for some $x \in A$, then*

$$\lim_{n\to\infty} n[B_n(f,x) - f(x)] = \frac{x(1-x)}{2} f''(x)\,. \tag{3.1}$$

If $f \in C^2(A)$, the convergence is uniform.

Historically, this has been a first example of saturation: for certain operators, convergence cannot be too fast, even for very smooth functions. The theorem shows that $f(x) - B_n(f,x)$ is of order not better than $1/n$ if $f''(x) \neq 0$.

Proof. We have

$$f\left(\frac{\nu}{n}\right) = f(x) + \left(\frac{\nu}{n}-x\right)f'(x) + \left(\frac{\nu}{n}-x\right)^2\left[\frac{1}{2}f''(x) + h\left(\frac{\nu}{n}-x\right)\right]\,, \tag{3.2}$$

where $h(y) := h_x(y)$ is bounded for all y and converges to zero with y. This yields

$$
\begin{aligned}
B_n(f,x) &= \sum_{\nu=0}^{n} f\left(\frac{\nu}{n}\right) p_{n,\nu}(x) \\
&= f(x) + \frac{1}{2} f''(x) T_{n,2}(x) + n^{-2} \sum_{\nu=0}^{n} (\nu - nx)^2 h\left(\frac{\nu}{n} - x\right) p_{n,\nu}(x) .
\end{aligned}
$$

By the remark (1.7), the last term does not exceed $o(1/n)$, and we obtain (3.1). If $f \in C^2(A)$, then $h_x(y) \to 0$ uniformly in $x \in A$ as $y \to 0$. □

As a corollary, we have the *localization theorem:*

if the functions $f_1, f_2 \in C(A)$ *coincide on a subinterval* $[a,b]$ *of* A, *then*

(3.3) $$B_n(f_1,x) - B_n(f_2,x) = o(1/n) , \quad x \in (a,b) .$$

If $f \in C^2(A)$, Taylor's fromula ((5.6) of Chapter 2) shows that the last term in (3.2) does not exceed $\frac{1}{2}\|f''\|(\frac{\nu}{n} - x)^2$. This leads to the inequality

(3.4) $$\|f - B_n(f)\| \le \|f''\| \frac{x(1-x)}{2n} , \qquad f \in C^2(A) .$$

Using the boundedness of the operator B_n and Theorem 5.3 of Chapter 7, we obtain

Theorem 3.2 (Popoviciu [1935], Freud [1968]). *For some constant* $C > 0$, *and all* $f \in C(A)$,

(3.5) $$|f(x) - B_n(f,x)| \le C\omega_2(f, \sqrt{X/n}) , \qquad x \in A .$$

In particular, if $f \in C^1(A)$, *then*

(3.6) $$|f(x) - B_n(f,x)| \le C\sqrt{X/n}\,\omega(f', \sqrt{X/n}) , \qquad x \in A .$$

The rate of convergence given by inequalities (3.5) and (3.6) is much worse than that given in Jackson's theorems of Chapter 7. This slow convergence of Bernstein polynomials is compensated for by their *shape preserving* properties.

Theorem 3.3. (i) *The polynomial* $B_n(f)$ *increases on* A *if* f *is increasing on this interval;*
(ii) *For* $k = 1, 2, \ldots$, $B_n(f)$ *is monotone of order* k *on* A *if* f *has this property;*
(iii) $\operatorname{Var} B_n(f) \le \operatorname{Var} f$*;*
(iv) *one has* $Z_{(0,1)} B_n(f) \le S_{(0,1)} f$ *where the first term is the number of zeros of* $B_n(f)$ *on* $(0,1)$ *and the second term is the number of sign changes of* f *on* $(0,1)$ (Pólya in Schoenberg [1959]).

Proof. The statements (i) and (ii) follow from (2.2) and (2.3) respectively. To prove (iii), we use (1.8):

$$
\begin{aligned}
\operatorname{Var} B_n(f) &= \int_0^1 |B_n'(f,x)|\,dx \le n \sum_{\nu=0}^{n-1} \left|\Delta f\left(\frac{\nu}{n}\right)\right| \int_0^1 p_{n-1,\nu}(x)\,dx \\
&= \sum_{\nu=0}^{n-1} \left| f\left(\frac{\nu}{n}\right) - f\left(\frac{\nu-1}{n}\right)\right| \le \operatorname{Var} f .
\end{aligned}
$$

To prove (iv), we need the Descartes' rule of signs. For a polynomial $P(u) = \sum_{k=0}^n b_k u^k$, this rule says that the number of zeros $Z_{(0,+\infty)}P$ of P on $(0,+\infty)$ counting multiplicities, does not exceed the number of changes of signs, $S(b_k)$, of the sequence $b_0, \ldots, b_n$, of the coefficients of P. Here, zeros in the sequence b_k are omitted. We have

$$B_n(f,x) = (1-x)^n \sum_{\nu=0}^{n} \binom{n}{\nu} f\left(\frac{\nu}{n}\right) u^\nu =: (1-x)^n Q(u) ,$$

where $u := \frac{x}{1-x}$. It is easy to check that $Z_{(0,1)}(B_n(f)) \le Z_{(0,+\infty)}(Q)$ (with multiplicities counted). The Descartes rule then gives

$$Z_{(0,1)}(B_n(f)) \le S\left\{f\left(\frac{\nu}{n}\right)\right\}_{\nu=1}^{n} \le S_{(0,1)} f . \qquad \square$$

Operators U on $C[a,b]$ which satisfy (iv) are called *nullity diminishing*, and those satisfying (iii) are *variation diminishing.* A useful remark is the following:

Remark. A nullity diminishing operator U which preserves constants is variation diminishing.

The proof follows at once from Banach's indicatrix formula $Var_{(a,b)} f = \int_{-\infty}^{\infty} N(f,\lambda)\,d\lambda$, for the variation of a function $f \in C[a,b]$. Here $N(f,\lambda)$, for each $\lambda \in \mathbb{R}$, is the number of solutions of the equation $f(x) = \lambda$. Indeed, we then have $N(Uf,\lambda) \le N(f,\lambda)$ for all λ.

§ 4. Bernstein Polynomials of Convex Functions

We begin with the following formula of Averbach (given in Karlin [B-1968], p. 306) which contains divided differences of the second order.

Theorem 4.1. *For* $n = 1, 2, \ldots$, *we have*

(4.1) $$B_n(f,x) - B_{n+1}(f,x) = \frac{x(1-x)}{n(n+1)} \sum_{k=0}^{n-1} \left\{ \left[\frac{k}{n}, \frac{k+1}{n+1}, \frac{k+1}{n}\right] f \right\} p_{n-1,k}(x) .$$

Proof. We have

$$
\begin{aligned}
B_{n+1}(f,x) &= f(0)(1-x)^{n+1} + f(1)x^{n+1} \\
&\quad + \sum_{\nu=1}^{n} f\left(\frac{\nu}{n+1}\right)\binom{n+1}{\nu} x^{\nu}(1-x)^{n+1-\nu}, \\
B_n(f,x) &= B_n(f,x)(x+(1-x)) = f(0)(1-x)^{n+1} + f(1)x^{n+1} \\
&\quad + \sum_{\nu=1}^{n} f\left(\frac{\nu-1}{n}\right)\binom{n}{\nu-1} x^{\nu}(1-x)^{n+1-\nu} \\
&\quad + \sum_{\nu=1}^{n} f\left(\frac{\nu}{n}\right)\binom{n}{\nu} x^{\nu}(1-x)^{n+1-\nu}.
\end{aligned}
$$

Subtracting and replacing $\nu - 1$ by k in all sums, we obtain

$$
\begin{aligned}
B_n(f,x) - B_{n+1}(f,x) &= x(1-x)\sum_{k=0}^{n-1} x^k(1-x)^{n-k} \\
&\quad \times \left[f\left(\frac{k}{n}\right)\binom{n}{k} - f\left(\frac{k+1}{n+1}\right)\binom{n+1}{k+1} + f\left(\frac{k+1}{n}\right)\binom{n}{k+1}\right].
\end{aligned}
\tag{4.2}
$$

By formula (7.7) of Chapter 4 for second divided differences,

$$
\begin{aligned}
\left[\frac{k}{n}, \frac{k+1}{n+1}, \frac{k+1}{n}\right] f &= f\left(\frac{k}{n}\right)\frac{n^2(n+1)}{(n-k)} - f\left(\frac{k+1}{n+1}\right)\frac{n^2(n+1)}{(k+1)(n-k)} \\
&\quad + f\left(\frac{k+1}{n}\right)\frac{n^2(n+1)}{k+1},
\end{aligned}
$$

hence (4.2) is equivalent to (4.1). □

If f is convex, then all terms in the sum (4.1) are non-negative and we obtain:

Corollary 4.2 (Schoenberg [1959], Arama [1960]). *If f is convex on A, then*

$$
B_n(f,x) \geq B_{n+1}(f,x) \geq f(x), \qquad 0 < x < 1. \tag{4.3}
$$

The inequalities are strict if f is strictly convex on A.

One can invert this statement.

Theorem 4.3. *If $f \in C[a,b]$, $0 \leq a < b \leq 1$, is bounded on A, and if*

$$
\limsup\{nB_n(f,x) - f(x)]\} \geq 0, \qquad a < x < b, \tag{4.4}
$$

then f is convex on $[a,b]$.

Proof. Suppose that f is not convex. Then for some $\varepsilon > 0$, neither is $g(x) = f(x) + \varepsilon x^2$. Thus, there are points $x_1 < y < x_2$ in $[a,b]$ for which $g(y) > l(y)$

where l is the linear function interpolating g: $l(x_i) = g(x_i)$, $i = 1,2$. Now $g - l$ is continuous on $I := [x_1, x_2]$. Let $m > 0$ be the maximum of $g - l$ on I, and suppose it is assumed at z with $x_1 < z < x_2$. Then $f(x) \leq -\varepsilon x^2 + l(x) + m$ with equality at z. Hence, for some constant α,

$$
f(x) \leq -\varepsilon(x-z)^2 + \alpha(x-z) + f(z) =: Q(x), \qquad x \in I.
$$

We can extend f to a function $F \in C(A)$, so that

$$
\begin{aligned}
F(x) &= f(x), \qquad x \in I, \\
F(x) &\leq Q(x), \qquad x \in A.
\end{aligned}
$$

Since $B_n(u^2, x) = x^2 + Xn^{-1}$, we have

$$
n[B_n(F,z) - F(z)] \leq n[B_n(Q,z) - Q(z)] = -\varepsilon z(1-z).
$$

The localization property (3.3) then shows that

$$
n[B_n(f,z) - f(z)] \leq -\varepsilon z(1-z) + o(1),
$$

which contradicts the assumption (4.4). □

§ 5. Saturation and Inverse Theorems

As with all positive operators, the Bernstein polynomials possess saturation. No function $f \in C(A)$, $A := [0,1]$ can be approximated with error better than $o(1/n)$ unless it is linear. But unlike saturation theorems for $\mathbb{T}$ (see §2 of Chapter 11) best estimates for $B_n(f)$ are *local* in character, that is, they depend on x. The following theorem of Lorentz [1964] admits a simple proof based on Theorem 4.3 and convexity.

The space $W^2_\infty(A)$ consists of all functions f on A for which f' is absolutely continuous and $|f''| \leq M$, a.e. for some M. By Theorem 9.3 of Chapter 2, W^2_∞ is equivalent with $\mathrm{Lip}(2, L_\infty)$ and the last inequality is equivalent to $\omega_2(f,h) \leq Mh^2$.

Theorem 5.1. *The saturation order of the Bernstein polynomials is X/n, and the saturation class is W^2_∞. More precisely, for each $M \geq 0$,*

$$
|f(x) - B_n(f,x)| \leq M\frac{X}{2n} + o_x\left(\frac{1}{n}\right), \qquad x \in A,\ n = 1,2,\dots \tag{5.1}
$$

holds exactly when $f \in W^2_\infty$ and $\|f''\|_\infty \leq M$. Relation (5.1) is equivalent to

$$
|f(x) - B_n(f,x)| \leq M\frac{X}{2n}, \qquad x \in A,\ n = 1,2,\dots. \tag{5.2}
$$

Here, $o_x(1/n)$ signifies any function $\varepsilon_n(x)$ for which $n\varepsilon_n(x) \to 0$ for each x. *In particular, $f(x) - B_n(f,x) = o_x(1/n)$ if and only if f is linear* (see also Bajsanski and Bojanic [1964]).

Proof. Let a function $f \in W^2_\infty(A)$ with $|f''(x)| \le M$, a.e., and a point $x \in A$ be fixed. For the linear function $l(y) := f(x)+f'(x)(y-x)$, by Taylor's formula, we have

$$|f(y) - l(y)| \le \tfrac{1}{2}M(y-x)^2 .$$

Since B_n preserves linear functions, $B_n(l,x) = l(x)$, hence

$$|f(x) - B_n(f,x)| = |B_n(f-l,x)| \le \frac{1}{2}MB_n((\cdot - x)^2, x) = M\frac{X}{2n} .$$

Conversely, if (5.1) is satisfied, we put $g(u) := f(u) + \frac{1}{2}Mu^2$ and obtain

$$n\left[B_n(g,x) - g(x)\right] \ge o_x(1/n) .$$

Hence, (4.4) is satisfied for g on A. Thus, g is convex and therefore

$$g(x+t) - g(x-t) - 2g(x) \ge 0 .$$

This means that

$$f(x+t) + f(x-t) - 2f(x) \ge -Mt^2 .$$

Similarly, taking $g(u) := -f(u) + \frac{1}{2}Mu^2$, we find that

$$f(x+t) + f(x-t) - 2f(x) \le Mt^2 ,$$

so that indeed, $|\Delta_t^2(f,x)| \le Mt^2$. □

More generally, if $0 < \alpha \le 2$, and if the function $f \in C(A)$ satisfies

$$\omega_2(f,t) \le Mt^\alpha , \quad t \ge 0 , \quad \text{for some } M \ge 0 , \tag{5.3}$$

then $|f(x) - B_n(f,x)| \le CM(X/n)^{\alpha/2}$ with some absolute constant $C > 0$. This follows from Theorem 3.2.

We shall prove the converse of this statement, see Theorem 5.3 below. In other words, we shall describe the functions corresponding to the approximation

$$|f(x) - B_n(f,x)| \le M(X/n)^{\alpha/2} , \qquad x \in A , \ n = 1, 2, \dots . \tag{5.4}$$

We need the following key lemma

Lemma 5.2 (Berens and Lorentz [1972]). *If $0 < \alpha < 2$, if ϕ is an increasing positive function on $[0,a]$ with $\phi(0) = 0$, and if $0 < a \le 1$, then the inequalities*

$$\phi(a) \le M_0 a^\alpha \tag{5.5}$$

and

$$\phi(x) \le M_0\left(y^\alpha + \frac{x^2}{y^2}\phi(y)\right) , \qquad 0 \le x \le y \le a , \tag{5.6}$$

imply for some $C = C(\alpha) > 0$,

$$\phi(x) \le CM_0x^\alpha , \qquad 0 \le x \le a . \tag{5.7}$$

Proof. For $0 < q < 1$, we define $x_k := q^k a$, $k = 0, 1, \dots$. If we take $C \ge 1$, then (5.5) implies (5.7) for x_0. We prove (5.7) for all $x = x_k$ by induction. Let $\phi(x_k) \le CM_0x_k^\alpha$, then from (5.6),

$$\phi(x_{k+1}) \le M_0\left(x_k^\alpha + q^2\phi(x_k)\right) \le M_0\left(1 + q^2CM_0\right)x_k^\alpha \le M_0Cx_{k+1}^\alpha ,$$

provided $1 + q^2CM_0 \le Cq^\alpha$. To achieve this, we first take q so small that $q^\alpha > M_0q^2$, and then C sufficiently large. After this, for any $0 < x < a$, we select a k with $x_{k+1} \le x \le x_k$ and get with $C_1 := Cq^{-\alpha}$,

$$\phi(x) \le \phi(x_k) \le CM_0x_k^\alpha \le C_1M_0x^\alpha .$$ □

The key to the inverse theorem lies in two different estimates of $B_n''(f)$. The identity (2.3), with $k = 2$, shows that

$$|B_n''(f,x)| \le n(n-1) \max_{0 \le \nu \le n-2}\left|\Delta^2\left(f\left(\frac{\nu}{n}\right)\right)\right| \le n^2\omega_2\left(f, \frac{1}{n}\right) \le \|f''\|_\infty . \tag{5.8}$$

We can also continue along the lines of (2.2) and derive the formula

$$B_n''(f,x) = X^{-2}\sum_{\nu=0}^n f\left(\frac{\nu}{n}\right)\left[(\nu - nx)^2 - (1-2x)\nu - nx^2\right]p_{n,\nu}(x) ,$$

from which it follows that

$$|B_n''(f,x)| \le X^{-2}\|f\|_\infty \sum_{\nu=0}^n \left|(\nu - nx)^2 - (1-2x)\nu - nx^2\right| p_{n,\nu}(x) . \tag{5.9}$$

The function $(1-2x)\nu + nx^2$, whose minimum is attained at $x = \nu/n$, is ≥ 0 on A. Consequently, the sum in (5.9) does not exceed

$$\begin{aligned}\sum_{\nu=0}^n \{(\nu - nx)^2 + (1-2x)\nu + nx^2\}p_{n,\nu}(x) &= T_{n,2}(x) + (1-2x)T_{n,1}(x) + nX \\ &= 2nX .\end{aligned}$$

Thus,

$$|B_n''(f,x)| \le \frac{2n}{X}\|f\|_\infty . \tag{5.10}$$

This inequality can be improved at the endpoints. From (5.8), it also follows that $|B_n''(f,x)| \le 4n^2\|f\|_\infty$. Hence,

$$|B_n''(f,x)| \le 4n\min(n, X^{-1})\|f\|_\infty . \tag{5.11}$$

Theorem 5.3 (Berens and Lorentz [1972]). *Let $0 < \alpha < 2$. There exists a constant $C = C(\alpha) > 0$ such that whenever $f \in C(A)$ and*

$$|f(x) - B_n(f,x)| \le M\left(\frac{X}{n}\right)^{\alpha/2} , \qquad x \in A , \ n = 1, 2, \dots , \tag{5.12}$$

then $\omega_2(f,h) \le CMh^\alpha$, $h > 0$.

Proof. The main idea of the proof is to use the equivalence of the modulus of smoothness $\omega_2(f,t)$ with the K-functional $K(f,t^2) := K(f,t^2;C,C^2)$ (Theorem 2.4 of Chapter 6). Without loss of generality, let $M=1$. For the function $\phi(t) := \omega_2(f,t)$, which is defined on $[0,1/2]$, we shall establish the hypotheses of Lemma 5.2 for $a=1/2$. Since $B_1(f)$ is a linear function, (5.12) gives that

$$\omega_2(f,a) = \omega_2(f-B_1(f),a) \le 4\|f-B_1(f)\|_\infty \le 4\,,$$

and therefore (5.5) is satisfied for any $M_0 \ge 2^{2+\alpha}$.

To prove (5.6), it is enough to show that for all x, $0 \le x \le 1-2h$,

$$(5.13)\qquad |\Delta_h^2(f,x)| \le M_0\left[\delta^\alpha + \frac{h^2}{\delta^2}\omega_2(f,\delta)\right], \qquad 0\le h\le\delta\le a$$

for M_0 sufficiently large. We shall prove (5.13) when $1/2 \le x \le 1-2h$; the case $0\le x\le 1/2$ is treated similarly.

For the prescribed values of x and δ, let $J := [x,x+2\delta]\cap A$. If $0\le h\le\delta$ and $x+2h\le 1$, then

$$(5.14)\qquad \begin{aligned}|\Delta_h^2(f,x)| &\le |\Delta_h^2(f-B_n(f),x)| + |\Delta_h^2(B_n(f),x)|\\ &\le 4\|f-B_n(f)\|_\infty(J) + h^2\|B_n''(f)\|_\infty(J)\,.\end{aligned}$$

For the first term on the right of (5.14), we note that the function X of x decreases for $1/2\le x\le 1$. For our fixed x, we take $n \ge \frac{1}{2}\delta^{-2}X$ and obtain from (5.12)

$$(5.15)\qquad \|f-B_n(f)\|_\infty(J) \le \left(\frac{X}{n}\right)^{\alpha/2} \le \left(2\delta^2\right)^{\alpha/2} \le 2\delta^\alpha\,.$$

Let z be the right endpoint of J. To estimate the second term on the right of (5.14), we need a value of n that satisfies in addition

$$(5.16)\qquad n\min(n,Z^{-1}) \le C_1\delta^{-2}\,.$$

We select n in different ways if δ or if X is the larger. We put $A := \max(3\delta^{-1},\delta^{-2}X)$ and define the positive integer n by

$$(5.17)\qquad n\le A<2n\,.$$

For this n, we have (5.15). Moreover, if $A=3\delta^{-1}$ then (5.16) follows from $n\le A\le 3\delta^{-1}$. On the other hand, if $A=\delta^{-2}X$, then $X\ge 3\delta$ and $Z\ge X-2\delta\ge 3^{-1}X$. Then, $nZ^{-1}\le 3AX^{-1} = 3\delta^{-2}$. Therefore, (5.16) is valid when $C_1 = 9$.

According to (3.10) of Chapter 6, with $r=2$ and $p=\infty$, there exists a function $f_\delta\in C^2(A)$ with the property

$$(5.18)\qquad \|f-f_\delta\|_\infty + \delta^2\|f_\delta''\|_\infty \le C_0\omega_2(f,\delta)\,.$$

We can estimate the second term on the right in (5.14) by means of $B_n''(f) = B_n''(f-f_\delta) + B_n''(f_\delta)$. By using (5.11) and (5.8), respectively, and afterwords (5.18) and (5.16), we obtain

$$\begin{aligned}\|B_n''(f)\|_\infty(J) &\le 4n\min(n,Z^{-1})\|f-f_\delta\|_\infty + \|f_\delta''\|_\infty\\ &\le C_0\omega_2(f,\delta)(n\min(nZ^{-1})+\delta^{-2}) \le C_0\delta^{-2}\omega_2(f,\delta).\end{aligned}$$

This establishes (5.13). □

Besides the proof for Theorem 5.3 given above, and the original proof, there are quite different proofs of DeVore in [A-1972] and Becker [1979].

It is worthwhile to compare the approximation spaces corresponding to $|f(x)-B_n(f,x)| = \mathcal{O}((X/n)^{\alpha/2})$ with the Lipschitz spaces of §9, Chapter 2. For $0<\alpha<2$, $\alpha\ne 1$, the condition $\omega_2(f,t) = \mathcal{O}(t^\alpha)$ is equivalent to $f\in \mathrm{Lip}(\alpha,\infty)$, by Theorem 7.2; for $\alpha=1$, this is not Lip 1, but the Zygmund space Z. Therefore, we have

Corollary 5.4. *The approximation space for the relation $|f(x)-B_n(f,x)| = \mathcal{O}((X/n)^{\alpha/2})$, $0<\alpha\le 2$ is the space $\mathrm{Lip}(\alpha,\infty)$, with the exception of the case $\alpha=1$, when it is the space Z.*

§ 6. Saturation Theorems for Kantorovich Polynomials

Bernstein polynomials are not defined for integrable functions. For the Kantorovich polynomials $K_n(f)$ (see §1) however, interesting saturation problems arise for the spaces L_p. They have been settled by Maier [1978] and Riemenschneider [1978].

The saturation order here is again $1/n$. The polynomials $K_n(f)$ have the representation $K_n(f,x) = \int_0^1 L_n(x,t)f(t)\,dt$ with the kernel

$$(6.1)\qquad L_n(x,t) := \sum_{k=0}^n p_{n,k}(x)\,(n+1)\chi_{I_k}(t)\,, \qquad I_k := \left(\frac{k}{n+1},\frac{k+1}{n+1}\right).$$

Since $\int_0^1 L_n(x,t)\,dx = \int_0^1 L_n(x,t)\,dt = 1$, K_n maps each rearrangement invariant Banach function space, in particular each space L_p, $1\le p\le\infty$, into itself with norm one. From Theorem 4.4 of Chapter 2, it follows that $\|f-K_n(f)\|_p\to 0$, $n\to\infty$, for each $f\in L_p(A)$, $1\le p<\infty$. Indeed, for continuous f, which are dense in L_p, we have the uniform convergence $K_n(f)\to f$, since then $\|K_n(f)-B_n(f)\|_\infty\to 0$.

The "direct" statements of the saturation theorems depend upon the values of K_n for the logarithmic functions $g_1(x) := \log x$, $g_2(x) := \log(1-x)$, and $g(x) := \log x - \log(1-x)$. The importance of the latter function lies in the fact that is an indefinite integral of $1/X$.

Lemma 6.1 (Maier's inequality). *For some absolute constant C,*

$$(6.2)\qquad |K_n(g_1,x)-g_1(x)| \le \sum_{n+1}^\infty \frac{(1-x)^j}{j} + C\sum_{k=0}^n\left[\frac{1}{(k+1)^2}+\frac{1}{n}\right]p_{n,k}(x)\,.$$

Proof. We have

$$K_n(g_1,x) = \sum_{k=0}^{n+1}(n+1)\int_{I_k}\log u\,du\; p_{n,k}(x)$$
$$= \sum_{k=0}^{n}\log\left(e^{-1}\left(1+\frac{1}{k}\right)^k\frac{k+1}{n+1}\right)p_{n,k}(x)\,.$$

Also, from (1.9),

$$\log x = \log[1-(1-x)]$$
$$= -\sum_{j=1}^{\infty}\frac{(1-x)^j}{j} = \sum_{k=0}^{n}(S_k-S_n)p_{n,k}(x) - \sum_{j=n+1}^{\infty}\frac{(1-x)^j}{j}\,.$$

Hence,

$$|K_n(g_1,x)-g_1(x)|$$
$$\le \sum_{j=n+1}^{\infty}\frac{(1-x)^j}{j} + \sum_{k=0}^{n}\left|\log\frac{k+1}{n+1}+S_n-S_k+k\log\left(1+\frac{1}{k}\right)-1\right|p_{n,k}(x)\,.$$

Here, for $k\ge 1$,

$$k\log\left(1+\frac{1}{k}\right)-1 = \sum_{j=2}^{\infty}\frac{(-1)^{j-1}}{j}k^{-j+1} = \frac{-1}{2k}+\mathcal{O}(k^{-2})\,.$$

Therefore,

$$S_k-\log(k+1) = \sum_{j=1}^{k}\left(\frac{1}{j}-\log\left(1+\frac{1}{j}\right)\right) = \sum_{j=1}^{k}\left(\frac{1}{2j^2}+\mathcal{O}(j^{-3})\right),$$

and we find

$$\log\frac{k+1}{n+1}+S_n-S_k = \sum_{j=k+1}^{n}\left(\frac{1}{2j^2}+\mathcal{O}(j^{-3})\right) = \sum_{j=k+1}^{n}\left(\frac{1}{2j(j+1)}+\mathcal{O}(j^{-3})\right)$$
$$\le \sum_{j=k+1}^{\infty}\frac{1}{2j(j+1)}+\mathcal{O}\left(\frac{1}{k^2}\right)+\mathcal{O}\left(\frac{1}{n}\right) = \frac{1}{2k}+\mathcal{O}\left(\frac{1}{k^2}\right)+\mathcal{O}\left(\frac{1}{n}\right).$$

This gives the required inequality. □

Lemma 6.2. *With $\phi(x)=X=x(1-x)$, $1\le p\le\infty$, and p' the conjugate of p, we have*

(6.3) $$\|\phi^{1/p'}(K_n(g)-g)\|_p = \mathcal{O}(1/n)\,,\qquad n=1,2,\dots\,.$$

The same estimate holds for the functions $g_1=\log x$ and $g_2=\log(1-x)$, with $\phi=x$ and $\phi=1-x$, respectively.

Proof. We consider only the function g_1; the other two cases are handled similarly. Since

$$\int_0^1(1-x)^j\,dx = \frac{1}{j+1}\,,\qquad \int_0^1 p_{n,k}(x)\,dx = \frac{1}{n+1}$$

from (6.2), we have

$$\int_0^1|K_n(g_1,x)-g_1(x)|\,dx = \mathcal{O}(1/n)\,,$$

which is (6.3) when $p=1$. When $p=\infty$, we note that the maximum of $x(1-x)^j$ on A is taken at $x=1/(j+1)$ and is $\mathcal{O}(1/(j+1))$. Therefore, $x\sum_{j=n+1}^{\infty}(1-x)^j/j=\mathcal{O}(1/n)$. Also, $\frac{1}{n}\sum_{k=0}^{n}xp_{n,k}(x)\le 1/n$. In the remaining sum to be estimated, we put $S_k(x):=\sum_{\nu=0}^{k}xp_{n,\nu}(x)$, $S_{-1}:=0$. Here, $xp_{n,\nu}$ has its maximum at $(\nu+1)/(n+1)$, so that S_k decreases for $x\ge(k+1)/(n+1)$ and $S_k(x)\le(k+1)/(n+1)$, $x\in A$. By partial summation,

$$\sum_{k=0}^{n}\frac{1}{(k+1)^2}xp_{n,k}(x)\le C\left\{\sum_{k=0}^{n-1}\frac{1}{(k+1)^3}S_k(x)+\frac{1}{(n+1)^2}S_n(x)\right\} = \mathcal{O}\left(1/n\right),$$

and we obtain (6.3) for $p=\infty$.

Finally, for $1<p<\infty$, we use the inequality

$$\|x^{1/p'}f(x)\|_p^p\le\|x^{p/p'}|f(x)|^{p/p'}\|_\infty\|f\|_1\,.$$ □

The saturation theorem for the K_n characterizes those functions $f\in L_p(A)$, $1\le p\le\infty$ for which

(6.4) $$\|K_n(f)-f\|_p = \mathcal{O}\left(1/n\right),\qquad n=1,2,\dots\,.$$

It contains the essential localization factor $X=x(1-x)$. Namely, (6.4) holds if and only if f is absolutely continuous and for $1<p\le\infty$, Xf' is an absolutely continuous function whose derivative is in L_p; for $p=1$, Xf' is of bounded variation on A. Somewhat more formally,

Theorem 6.3. (i) *For $f\in L_p(A)$, $1\le p\le\infty$, one has (6.4) if and only if*

(6.5) $$f(x) = Const. + \int_a^x\frac{h(u)}{U}\,du\,,\qquad a.e.\ x\in A$$

where $a\in(0,1)$ and $h(0)=h(1)=0$. For $1<p\le\infty$, h is absolutely continuous with $h'\in L_p$; for $p=1$, h is of bounded variation on A.
(ii) *Moreover, if $\|K_n(f)-f\|=o(1/n)$, then f is a.e. constant.*

The necessity part of this theorem is due to Maier, the sufficiency part to him for $p=1$ and to Riemenschneider for $1<p\le\infty$.

Lemma 6.4. *Let $g(u) := \log u - \log(1-u)$ and let h be of bounded variation on A. Then, with the operator K_n applied with respect to t,*

$$\left\| K_n\left(\int_t^x [g(u)-g(t)]\, dh(u), x\right)\right\|_1 \le \frac{C}{n}\operatorname{Var} h . \tag{6.6}$$

Proof. Let L_n be the kernel (6.1). The norm in (6.6) is equivalent to

$$\int_0^1 \left| K_n\left(\int_t^x, x\right)\right| dx \le \int_0^1\int_0^1 dx\, dt \left|\int_t^x F\, dh(u)\right| , \tag{6.7}$$

where $F := L_n(x,t)[g(u)-g(t)]$. Now the function g increases from $-\infty$ to $+\infty$ on A. Therefore $F > 0$ in the interior integral on the right of (6.7) if $t < x$, and $F < 0$, if $t > x$. If $h^*(u)$ is the total variation of $h(u)$, then the absolute value on the right side of (6.7) does not exceed

$$\int_t^x F\, dh^* = \int_t^1 F\, dh^*(u) - \int_x^1 F\, dh^*(u) ,$$

and therefore the left side of (6.7) does not exceed

$$\begin{aligned}&\le \int_0^1 dt \int_t^1 dh^*(u)\int_0^1 F\, dx \;-\; \int_0^1 dx\int_x^1 dh^*(u)\int_0^1 F\, dt\\ &= \int_0^1 dh^*(u)\left\{\int_0^u dt\int_0^1 F\, dx - \int_0^u dx\int_0^1 F\, dt\right\} .\end{aligned}$$

We finish the proof by showing that the expression Δ in the curly brackets is $\mathcal{O}(n^{-1})$, uniformly in u. We have,

$$\int_0^1 F\, dt = \int_0^1 L_n(x,t)[g(u)-g(t)]\, dt = g(u) - K_n(g,x) ,$$

and $\int_0^1 F\, dx = g(u) - g(t)$, so that $\Delta = \int_0^u [K_n(g,x)-g(x)]\, dx = \mathcal{O}(1/n)$ by (6.3) with $p=1$. □

The corresponding lemma for the space L_∞ is simpler:

Lemma 6.5. *For a bounded function h,*

$$\left\| K_n\left(\int_t^x [g(u)-g(t)]\, h(u)\, du, x\right)\right\|_\infty \le \frac{C}{n}\|h\|_\infty . \tag{6.8}$$

Proof. The absolute value of the integral does not exceed

$$\begin{aligned}&\|h\|_\infty \int_t^x [g(u)-g(t)]\, du\\ &\quad = \|h\|_\infty\{x(\log x - \log t) + (1-x)(\log(1-x) - \log(1-t))\} .\end{aligned}$$

Since the operator K_n is positive,

$$\begin{aligned}&\left| K_n\left(\int_t^x, x\right)\right|\\ &\quad\le \|h\|_\infty\{x[\log x - K_n(\log\cdot, x)] + (1-x)[\log(1-x) - K_n(\log(1-\cdot), x)]\}\end{aligned}$$

and we can apply (6.3) with $p = \infty$. □

(a) *Proof of sufficiency in Theorem* 6.3 (ii). Under the assumptions of the theorem, we prove that $f' \in L_p$, if $1 < p \le \infty$, and that

$$n\|K_n(f)-f\|_p \le \begin{cases} C\{\|f'\|_p + \|(Xf')'\|_p\} , & 1 < p \le \infty\\ C\{\|Xf'\|_\infty + \operatorname{Var}(Xf')\} , & p = 1 ,\end{cases} \tag{6.9}$$

establishing (6.4). The first statement follows from

$$f'(x) = \frac{h(x)}{X} = \frac{1}{x}\int_0^x h'(u)\, du - \frac{1}{1-x}\int_x^1 h'(u)\, du$$

and Hardy's Theorem 3.1 of Chapter 2.

We now prove (6.9) for $1 \le p \le \infty$. By partial integration for $x, t \in (0,1)$, we can represent $f(t) - f(x)$ in terms of Xf':

$$f(t) - f(x) = Xf'(x)[g(t) - g(x)] - \int_t^x [g(u)-g(t)] d(Uf'(u)) . \tag{6.10}$$

We apply the operator K_n to (6.10) with respect to the variable t and get

$$K_n(f,x) - f(x) = Xf'(x)[K_n(g,x) - g(x)] - K_n\left(\int_t^x, x\right) = G_1 - G_2 . \tag{6.11}$$

We begin by estimating the first term, G_1 in (6.11). If $1 < p \le \infty$, Lemma 6.2 yields $\|X(K_n g - g)\|_\infty = \mathcal{O}(n^{-1})$, so that $\|G_1\|_p \le Cn^{-1}\|f'\|_p$. If $p = 1$, then again by Lemma 6.3, $\|K_n g - g\|_1 \le Cn^{-1}$, and $\|G_1\|_1 \le Cn^{-1}\|Xf'\|_\infty$.

The second term G_2 can, for $p = 1$, be estimated directly by Lemma 6.4: $\|G_2\|_1 \le Cn^{-1}\operatorname{Var}(Xf')$. For other values of p we need the operator

$$S_n(F,x) := K_n\left(\int_t^x [g(u)-g(t)]\, F(u) du, x\right) , \qquad F \in L_1 .$$

For $p = 1$ and $p = \infty$, we get, from Lemmas 6.4 and 6.5, the estimates $\|S_n(F)\|_p \le \frac{C}{n}\|F\|_p$, $p = 1$ and $p = \infty$, respectively. From this, by the interpolation Theorem 4.4 of Chapter 2, this inequality follows for all $1 \le p \le \infty$: In particular, we get $\|G_2\|_p \le \frac{C}{n}\|(Xf')'\|_p$ for $1 < p \le \infty$. □

To prove the necessity in Theorem 6.3, we employ the bilinear functionals

$$A_n(f,\psi) := 2n\int_0^1 [K_n(f,x) - f(x)]\psi(x)\, dx , \qquad f \in L_p ,\ \psi \in C^2 . \tag{6.12}$$

Lemma 6.6. *For each fixed $\psi \in C^2(A)$ and $1 \le p \le \infty$, the functionals (6.12) have bounded norms on L_p:*

$$\|A_n(\cdot,\psi)\|_{L_p} \le C_\psi\,. \tag{6.13}$$

Moreover, there exists the limit

$$\lim_{n\to\infty} A_n(f,\psi) = \int_0^1 f(X\psi')'\,dx =: A(f,\psi)\,. \tag{6.14}$$

Proof. If $\psi \in C^2$, we have for some $\xi = \xi(u,t) \in (0,1)$,

$$\begin{aligned}\int_0^1 K_n(f)\psi\,dt &= \int_0^1 \sum_{k=0}^n p_{n,k}(t)(n+1)\\ &\quad\times \int_{I_k} f(u)[\psi(u) + (t-u)\psi'(u) + \frac{1}{2}(t-u)^2\psi''(\xi)]\,du\,dt\,.\end{aligned} \tag{6.15}$$

We decompose this into three integrals $J_j(n)$, $j = 1,2,3$, corresponding to the three terms in the square brackets. Since

$$(n+1)\int_0^1 tp_{n,k}(t)\,dt = \frac{k+1}{n+2}\,,\quad (n+1)\int_0^1 t^2p_{n,k}(t)\,dt = \frac{k+1}{n+2}\frac{k+2}{n+3}\,,$$

we get

$$\begin{aligned}|J_3(n)| &\le \|\psi''\|_\infty \sum_{k=1}^n \int_{I_k} f(u)\left(u^2 - 2\frac{k+1}{n+2}u + \frac{k+1}{n+2}\frac{k+2}{n+3}\right)du\\ &\le \frac{C}{n}\|\psi''\|_\infty\|f\|_1\,.\end{aligned}$$

In the same way, we get

$$J_2(n) \le \|\psi'\|_\infty\|f\|_1\,,\qquad J_1(n) = \int_0^1 f\psi\,du\,.$$

This yields

$$|A_n(f,\psi)| \le C\|f\|_1 \le C\|f\|_p\,,$$

with C depending on ψ.

To find the limit $A(f,\psi)$ of the functionals (6.12), we note that for the indefinite integral F of f, $\frac{d}{dx}B_{n+1}(F,x) = K_n(f,x)$, and $B_{n+1}(F,x)$ and $F(x)$ coincide for $x = 0,1$. Integrating (6.12) by parts,

$$A_n(f,\psi) = -2n\int_0^1 [B_{n+1}(F,x) - F(x)]\psi'(x)\,dx\,.$$

First, let $f \in C^1(A)$. Then by Theorem 3.1, $2n[B_{n+1}(F,x) - F(x)] \to Xf'(x)$ uniformly for $x \in A$, and we get $A(f,\psi) = -\int_0^1 Xf'\psi'\,dx$, in other words (6.14). This $A(f,\psi)$ is a linear functional on L_p. By the Banach-Steinhaus theorem, $\lim A_n(f,\psi)$ exists for all $f \in L_p(A)$ and is given by the integral $\int_0^1 f(X\psi')'\,dx$. □

(b) *Proof of the necessity in* Theorem 6.3 (i). Let $f \in L_p$ be fixed and let

$$\|K_n(f) - f\|_p = \mathcal{O}(n^{-1})\,,\qquad n = 1,2,\ldots\,. \tag{6.16}$$

We shall obtain a new integral representation for $A(f,\psi)$. First let $1 < p \le \infty$. The functions $2n(K_n(f) - f)$ are a bounded set in L_p, and by the weak compactness of the unit ball in this space, they possess a subsequence that is weakly convergent to some $h_1 \in L_p$. Then, $A(f,\psi) = \int_0^1 h_1\psi\,dx$. For our f, we get the relation, valid for all $\psi \in C^2(A)$,

$$\int_0^1 f(X\psi')'\,dx = \int_0^1 h_1\psi\,dx\,. \tag{6.17}$$

In particular, with $\psi = 1$, $\int_0^1 h_1\,dx = 0$.

The case $p = 1$ is similar. We need the fact that $2n[K_n(f,t) - f(t)]\,dt$ is a bounded sequence of measures in the dual space of $C(A)$. This yields

$$\int_0^1 f(X\psi')'\,dt = \int_0^1 \psi\,dh(x) \tag{6.18}$$

for some $h \in \mathrm{BV}$, $\int_0^1 dh = 0$, and we can assume that $h(0) = h(1) = 0$.

We would like to solve (6.17) or (6.18) for f. In the first place, the only solutions of the homogeneous equations

$$\int_0^1 g(X\psi')'\,dx = 0\,,\qquad \text{for all } \psi \in C^2 \tag{6.19}$$

are the constant functions. In fact, with $\psi = x$, we find $\int_0^1(1-2x)g\,dx = 0$. Hence, with $G(x) := \int_0^x(1-2t)g(t)\,dt$, we get from (6.19),

$$\int_0^1 [Xg(x) - G(x)]\psi''\,dx = 0\,.$$

Since $\psi'' \in C$ is arbitrary, we have $Xg(x) = G(x)$, a.e. Hence g is equivalent to an absolutely continuous function on $(0,1)$. Differentiation of the last relation shows that $g'(x) = 0$, a.e. or that $g(x) = $ const.

To prove (ii), we note that the integral H in (6.5) (with $h(x) := \int_0^x h_1\,dt$ in case $1 < p \le \infty$) also satisfies (6.17) or (6.18). Hence, $g := H - f$ is a solution to (6.19) and therefore is a constant. Moreover, if f satisfies (ii), then $h_1 = 0$ and $f = $ const. This completes the proof of Theorem 6.3. □

§ 7. Characterization of Approximation Spaces

In the previous sections, we have derived upper bounds for $f(x) - B_n(f,x)$ which depend on the position of x in $A := [0,1]$. Similar to §7 of Chapter 8, we would like now to obtain bounds for this approximation which are independent of x. We want to know, for example, which functions $f \in C(A)$ satisfy

$\|f - B_n(f)\| = \mathcal{O}(n^{-\alpha/2})$? Results of this type, will be in terms of K-functionals, and the Ditzian-Totik modulus of smoothness, that depend on the function $\varphi(x) := \sqrt{X} = \sqrt{x(1-x)}$. The function $\varphi(x) = \sqrt{x(1-x)}$ for $[0,1]$ replaces the function $\phi(x) = \sqrt{1-x^2}$ on $[-1,1]$ of Chapter 6. We shall treat only approximation in the uniform norm, $\|\cdot\| := \|\cdot\|_\infty(A)$, on A, in which case

$$K(f,t) := K(f,t;C,W_\infty^2(\varphi)) := \inf_{g\in W_\infty^2(\varphi)} \{\|f-g\|_\infty + t\|\varphi^2 g''\|_\infty\}, \tag{7.1}$$

$$\omega_2^\varphi(f,t) := \sup_{0\le h\le t} \|\tilde{\Delta}^2_{h\phi(\cdot)}(f,\cdot)\| \tag{7.2}$$

(see (6.1) and (6.6) of Chapter 6). Here, as before, we use the convention that the symmetric difference $\tilde{\Delta}_s^2(f,x) := f(x+s) - 2f(x) + f(x-s)$ is defined to be 0 whenever $[x-s, x+s]$ is not entirely contained in A. The space $W_\infty^2(\varphi)$ consists of all functions $g \in C(A)$ such that g' is absolutely continuous on A and $|g|_{W_\infty^2(\varphi)} := \|\varphi^2 g''\|_\infty$ is finite.

Recently, Totik [1991] proved the remarkable result

$$\|f - B_n(f)\| \sim \omega_2^\varphi(f, n^{-1/2}). \tag{7.3}$$

(with constants of equivalence independent of f.) We shall prove the direct (and easier) half of this inequality (see (7.21)) but only a much weaker form of the inverse inequality (see Theorem 7.5). Totik's result was preceeded by several other inverse estimates, most notably, the inequality

$$\omega_2^\varphi(f, n^{-1/2}) \le C(\|f - B_n(f)\| + \|f - B_{\lambda n}\|),$$

for any fixed integer $\lambda > 1$, of Ditzian and Ivanov [1991].

The following is the main result of this section.

Theorem 7.1 (Berens and Lorentz, Ditzian and Totik). *For $0 < \alpha < 2$, the following are equivalent:*

$$\begin{aligned} &\text{(i)} \quad \|f - B_n(f)\| = \mathcal{O}(n^{-\alpha/2}), \quad n = 1,2,\dots, \\ &\text{(ii)} \quad \omega_2^\varphi(f,t) = \mathcal{O}(t^\alpha), \quad t > 0. \end{aligned} \tag{7.4}$$

An essential tool in the proof of Theorem 7.1 is the equivalence (Ditzian [1980₁], Totik [1984]):

$$K(f,t^2) \sim \omega_2^\varphi(f,t), \quad t \ge 0, \tag{7.5}$$

which follows from Theorem 6.2 of Chapter 6. With this, Theorem 7.1 will follow from the equivalence

$$\|f - B_n(f)\| = \mathcal{O}(n^{-\alpha/2}) \iff K(f,t) = \mathcal{O}(t^\alpha). \tag{7.6}$$

due to Berens and Lorentz [1972]. (They proved this for a K-functional K_0 which is equivalent to K.)

Lemma 7.2. *If $g \in W_\infty^2(\varphi)$, then, with $M := \|g\|_{W_\infty^2(\varphi)} < \infty$,*

$$\|g - B_n(g)\| \le \frac{M}{2n}, \quad n = 1,2,\dots. \tag{7.7}$$

Proof. Let $g \in C^2(\varphi)$. If l denotes the linear Taylor interpolant to g at x, then

$$g(t) = l(t) + \int_x^t g''(s)(t-s)\,ds =: l(t) + R(g,t,x). \tag{7.8}$$

We first note that

$$|R(g,t,x)| \le M\frac{(t-x)^2}{\varphi(x)^2}, \quad x,t \in (0,1). \tag{7.9}$$

Indeed, if we multiply the integrand in (7.8) by $\varphi^2(s)/\varphi^2(s)$, we see that to establish (7.9), it is sufficient to show, for $x,t \in (0,1)$ and s between x and t,

$$\frac{|t-s|}{s(1-s)} \le \frac{|t-x|}{x(1-x)}. \tag{7.10}$$

If $x \le s \le t$, (7.10) follows from $\frac{t-s}{1-s} \le \frac{t-x}{1-x}$ and $\frac{1}{s} \le \frac{1}{x}$, and if $t \le s \le x$, it follows from $\frac{s-t}{s} \le \frac{x-t}{x}$ and $\frac{1}{1-s} \le \frac{1}{1-x}$.

Next, since $B_n(l) = l$, from (7.8) and (7.9), we have for $0 < x < 1$,

$$|g - B_n(g,x)| = |B_n(R(g,.,x),x)| \le M\varphi(x)^{-2}B_n((\cdot - x)^2, x)) = \frac{M}{2n}.$$

On the other hand, $B_n(g,x) = g(x)$ for $x = 0,1$. □

Lemma 7.3. *For $f \in C(A)$, we have*

$$\|f - B_n(f)\| \le 2\,K(f,n^{-1}), \quad n = 1,2,\dots. \tag{7.11}$$

Proof. From the definition of the K-functional, for each $\varepsilon > 0$ and $n = 1,2,\dots$, there is a function $g \in W_\infty^2(\varphi)$ such that

$$\|f-g\| + n^{-1}\|\varphi g''\| \le K(f,n^{-1}) + \varepsilon.$$

Therefore, using (7.7)

$$\begin{aligned} \|f - B_n(f)\| &\le \|f-g\| + \|g - B_n(g)\| + \|B_n(g-f)\| \le 2\|f-g\| + \frac{1}{2n}\|\varphi^2 g''\| \\ &\le 2\,K(f,n^{-1}) + 2\varepsilon. \end{aligned} \tag{7.12}$$

Letting $\varepsilon \to 0$, we obtain (7.11). □

There are inverse estimates to (7.11) which are similar to those of §5. To derive these, we need the following two inequalities for $B_n''(f)$.

Lemma 7.4. *For an absolute constant $C>0$, we have for each $n=1,2,\dots$*

$$\text{(7.13)}\qquad \begin{aligned} &\text{(i)} \quad \|\varphi^2 B_n''(f)\| \le Cn\|f\|, \quad f\in C(A),\\ &\text{(ii)} \quad \|\varphi^2 B_n''(g)\| \le C\|\varphi^2 g''\|, \quad g\in W_\infty^2(\varphi). \end{aligned}$$

Proof. The inequalilty (i) follows immediately from (5.10). To prove (ii), we can assume that $n\ge 2$ since otherwise $B_n''(g)=0$. We begin with the identity (2.3) for $B_n''(g)$. We can relate the differences appearing in (2.3) with divided differences. We use the representation (2.2) of Chapter 5 and obtain

$$\text{(7.14)}\qquad \begin{aligned} n(n-1)\Delta^2 g\left(\frac{\nu}{n}\right) &= 2\left(1-\frac1n\right)\left[\frac{\nu}{n},\frac{\nu+1}{n},\frac{\nu+2}{n}\right] g\\ &= \left(1-\frac1n\right)\int g''(u)M_\nu(u)\,du =: a_\nu \end{aligned}$$

where $M_\nu(u) := M(\frac{\nu}{n},\frac{\nu+1}{n},\frac{\nu+2}{n};u)$ is the piecewise linear B-spline. Using (7.14) in (2.3), we obtain

$$\text{(7.15)}\qquad |\varphi^2(x)(B_n(g)''(x)| \le \sum_{\nu=0}^{n-2}|a_\nu|\varphi^2(x)p_{n-2,\nu}(x).$$

We shall estimate the terms in the last sum. From $\int M_\nu\,du=1$, we find $M_\nu(\frac{\nu+1}{n})=n$. Hence, for $\nu=0$, $M_0(u)\le 2n^2u(1-u)$, $u\in[0,1]$, and with $M:=\|\varphi^2 g''\|$,

$$|a_0| \le 2n^2\int_0^{\frac2n}\varphi^2(u)|g''(u)|\,du \le 4Mn.$$

Since $x(1-x)^{n-1}$ has its maximum at $x=1/n$ on A, we have

$$\text{(7.16)}\qquad |a_0|\varphi^2(x)p_{n-2,0}(x) \le 4Mn(1-x)^{n-1}x \le 4Mn\left(1-\frac1n\right)^{n-1}\frac1n \le 4M.$$

A similar argument gives the same estimate in the case $\nu=n-2$.

For $1\le\nu<n-2$, and $u\in[\frac{\nu}{n},\frac{\nu+2}{n}]$, we have $\varphi(\frac{\nu+1}{n})\le\sqrt2\varphi(u)$. Therefore,

$$|a_\nu| \le 2\left[\varphi\left(\frac{\nu+1}{n}\right)\right]^{-2}\int_0^1\varphi^2(u)|g''(u)|M_\nu(u)\,du \le 2M\left[\varphi\left(\frac{\nu+1}{n}\right)\right]^{-2}$$

because $\int M_\nu\,du=1$. This gives

$$\text{(7.17)}\qquad \begin{aligned} |a_\nu|\varphi^2(x)p_{n-2,\nu}(x) &\le 2M\frac{n^2(n-2)!}{(\nu+1)!(n-\nu-1)!}x^{\nu+1}(1-x)^{n-\nu-1}\\ &\le 4Mp_{n,\nu+1}(x). \end{aligned}$$

The inequalities (7.16) and (7.17) when used in (7.15) yield

$$|\varphi^2(x)B_n''(g)(x)| \le 8M+4M\sum_{\nu=1}^{n-3}p_{n,\nu+1}(x) \le 12M.$$

□

As an inverse to (7.11), we have

Theorem 7.5. *For each $k,n=1,2,\dots$, we have*

$$\text{(7.18)}\qquad K(f,n^{-1}) \le C\left\{\|f-B_k(f)\|+\frac{k}{n}K(f,k^{-1})\right\}$$

where C is an absolute constant.

Proof. From the definition of the K-functional,

$$\text{(7.19)}\qquad K(f,n^{-1}) \le \|f-B_k(f)\|+n^{-1}\|\varphi^2B_k''(f)\|.$$

Given an $\varepsilon>0$ and a positive integer k, we choose $g\in W_\infty^2(\varphi)$ to satisfy

$$\|f-g\|+k^{-1}\|\varphi^2 g''\| \le K(f,k^{-1})+\varepsilon$$

and write $B_k''(f)=B_k''(f-g)+B_k''(g)$. By virtue of (7.13) (i) and (ii), we obtain

$$\begin{aligned} \|\varphi^2B_k''(f)\| &\le \|\varphi^2B_k''(f-g)\|+\|\varphi^2B_k''(g)\|\\ &\le Ck\{\|f-g\|+k^{-1}\|\varphi^2g''\|\} \le Ck\{K(f,k^{-1})+\varepsilon\}. \end{aligned}$$

When this is used in (7.19) and we let $\varepsilon\to0$, we obtain (7.18). □

Proof of Theorem 7.1. If (7.4) (ii) is satisfied, we deduce (7.4) (i) from Lemma 7.3 and the equivalence (7.5). The reverse implication follows from Lemma 5.2 and Theorem 7.5. Indeed, if (7.4) (i) is valid, we put $\phi(t):=K(f,t^2)$, $a=1$. Then (7.18) and the monotonicity of K give

$$\text{(7.20)}\qquad \phi(x)\le C(y^\alpha+y^2x^{-2}\phi(y)),\quad 0\le x\le y\le a.$$

Condition (5.5) is trivially satisfied and we deduce $\phi(t)=\mathcal{O}(t^\alpha)$. From the equivalence (7.5) of K and ω_2^φ, we obtain (ii). □

Corollary 7.6. *For each $f\in C(A)$, there holds the Jackson type inequality*

$$\text{(7.21)}\qquad \|f-B_n(f)\| \le C\omega_2^\varphi(f,n^{-1}).$$

with $C>0$ an absolute constant.

Corollary 7.7 (Ivanov[1982]). *For $f\in C(A)$, the following statement is equivalent to* (7.4) (i) *and* (ii)*:*

$$E_n(f)=\mathcal{O}(n^{-\alpha}).$$

The remaining (saturation) case $\alpha=2$ of Theorem 7.1 is also true.

Theorem 7.8. (i) *A function $f\in C(A)$ satisfies*

$$\text{(7.22)}\qquad \|f-B_n(f)\|_\infty \le \frac{M}{2n}+o\left(\frac1n\right),\qquad n=1,2,\dots$$

if and only if $f \in W^2_\infty(\varphi)$ *and* $\|\varphi^2 f''\| \le M$. (ii) *Moreover, relation* (7.22) *is equivalent to*

$$\|f - B_n(f)\|_\infty \le \frac{M}{2n}, \qquad n = 1, 2, \ldots . \tag{7.23}$$

In particular, $\|f - B_n(f)\| = o(n^{-1})$ *if and only if* f *is a linear function on* A.

The proof of (i) and (ii) is similar to that of Theorem 5.1 and is omitted here. (See Lorentz and Schmaker [1972], where (i) has been proved for many sequences of operators, among them, the Bernstein operators.)

We have only treated the Bernstein operator and approximation in the uniform norm. In Ditzian and Totik [A-1987], the authors discuss several related operators (of Kantorovich, Szász, Baskakov-Kantorovich, and others) and treat them in the L_p norm for $1 \le p \le \infty$ and with weighted moduli of smoothness of orders $r \ge 2$.

§ 8. Further Properties and Variants of Bernstein Polynomials. Weak Haar Spaces

The Bernstein polynomials $B_n(f, z)$ of an analytic function f have interesting properties. We mention the following simple result.

Theorem 8.1. *For an entire function* f, *the Bernstein polynomials* $B_n(f)$ *converge to* f *uniformly in any bounded region.*

Proof. In (2.4), we can replace the difference $\Delta^k f(0)$ with step $h = n^{-1}$ by $f^{(k)}(a_k)h^k$ for some a_k, $0 < a_k \le k/n$, and obtain

$$B_n(f, z) = \sum_{k=0}^{n} (1 - \frac{1}{n}) \ldots (1 - \frac{k-1}{n}) \frac{1}{k!} f^{(k)}(a_k) z^k . \tag{8.1}$$

For an arbitrary $r > 1$, we select $r_1 > r$ and then $M > 0$ so large that $|f(z)| \le M$, $|z - a| \le r_1$, for each a, $0 < a \le 1$. By Cauchy's inequality for the coefficients of the Taylor series of f at a_k, we obtain $\frac{1}{k!}|f^{(k)}(a_k)| \le M r_1^{-k}$. This shows that for $|z| \le r$, the finite series (8.1), $n = 1, 2, \ldots$, possess a common majorant, $M \sum_{k=0}^{\infty} (r/r_1)^k$. Therefore, we may pass to the limit in (8.1) termwise when $n \to \infty$ and obtain, $\lim B_n(f, z) = \sum_{k=0}^{\infty} f^{(k)}(0) z^k / k! = f(z)$, uniformly in $|z| \le r$. □

For a domain $G \supset [0, 1]$, there exists a subset G_0 with $G \supset G_0 \supset [0, 1]$ such that $B_n(f) \to f$ uniformly on G_0 for all f analytic in G. The dependence of G_0 on G is a deep problem, studied by Bernstein and developed further by one of us (Lorentz [A1986, Chapter IV]).

We shall discuss an interesting modification of the Bernstein (and the Kantorovich) polynomials that are obtained by replacing $f(k/n)$ in the classical formula by the moment of f with respect to $p_{n,k}$:

$$M_n(f) := M_n(f, x) := (n + 1) \sum_{k=0}^{n} p_{n,k}(x) \int_0^1 f(t) p_{n,k}(t)\, dt . \tag{8.2}$$

Introduced by Durrmeyer, they have been studied by Derrienic [1981], Ditzian and Ivanov [1991], and others.

As an example, we compute M_n for the functions $\phi_m := x^m$, $m = 0. \ldots, n$. It is clear that $M_n(1) = 1$. As in (1.8), we calculate

$$\int_0^1 p_{n,k}(t) t^m\, dt = \binom{n}{k} \int_0^1 t^{m+k} (1 - t)^{n-k}\, dt = \frac{(m + k)! n!}{k! (n + m + 1)!} .$$

Hence,

$$M_n(\phi_m, x) = \frac{(n + 1)!}{(n + m + 1)!} \sum_{k=0}^{n} \frac{(m + k)!}{k!} p_{n,k}(x) .$$

For the quantity $\dfrac{\partial^m}{\partial x^m}(x^m (x + y)^n)$, we can give two different expressions, one obtained by expanding $(x + y)^m$ and differentiating, the other obtained from Leibniz' formula. We obtain in this way the identity

$$\sum_{k=0}^{n} \binom{n}{k} x^k y^{n-k} \frac{(m + k)!}{k!} = \sum_{\nu=0}^{m} \binom{m}{\nu} \frac{m!}{\nu!} x^\nu \frac{n!}{(n - \nu)!} (x + y)^{n-\nu} .$$

For $y = 1 - x$, this yields the formula

$$M_n(\phi_m, x) = \frac{(n + 1)!}{(n + m + 1)!} \sum_{\nu=0}^{m} \binom{m}{\nu} \frac{m!}{\nu!} \frac{n!}{(n - \nu)!} x^\nu \tag{8.3}$$

This is a polynomial of degree m. For example,

$$M_n(\phi_1, x) = \frac{1 + nx}{n + 2}, \quad M_n(\phi_2, x) = \frac{2 + 4nx + n(n - 1)x^2}{(n + 2)(n + 3)} .$$

For the operators M_n, we can estimate the quantities λ_n of (4.1), Chapter 9. We easily obtain $\lambda_n \le 3(n + 2)^{-1}$.

Theorem 8.2. (i) *The operators* M_n *are positive and norm diminishing on each rearrangement invariant Banach function space* X *on* $A = [0, 1]$ *(for example, on* L_p, $1 \le p \le \infty$*).* (ii) *If* P *is a polynomial of degree* m, *then* $M_n(P)$ *is a polynomial of degree* $\le \min(m, n)$. (iii) *For* $f \in C(A)$, *we have*

$$\|f - M_n(f)\| \le 3\omega\left(f, \frac{\sqrt{3}}{\sqrt{n}}\right), \qquad n = 1, 2 \ldots . \tag{8.4}$$

Proof. We can represent $M_n(f)$ by means of an integral: $M_n(f,x) = \int_0^1 K_n(x,t)f(t)\,dt$, where K_n is the positive symmetric kernel $K_n(x,t) := (n+1)\sum_{k=0}^n p_{n,k}(x)p_{n,k}(t)$. It satisfies $\int_0^1 K_n\,dt = \int_0^1 K_n\,dx = 1$ and (i) follows from Theorem 4.5 of Chapter 2. From (8.3), we deduce (ii) while (iii) follows from Theorem 4.2 of Chapter 9. □

The following properties distinguish the M_n from the ordinary Bernstein polynomials.

Theorem 8.3. *The operators M_n are self adjoint on $L_2(A)$. Their eigenfunctions are the Legendre polynomials Q_k, $k = 0,\dots,n$, for the interval A, and the eigenvalues are*

$$\lambda_{n,k} = \frac{(n+1)!n!}{(n-k)!(n+k+1)!}, \qquad k = 0,\dots,n. \tag{8.5}$$

Proof. The first statement follows from the relation, valid for $f,g \in L_1(A)$,

$$(M_n(f),g) = \int_0^1 M_n(f,x)g(x)\,dx = \int_0^1 f(t)M_n(g,t)\,dt = (f, M_n(g)). \tag{8.6}$$

The orthonormal polynomials Q_k with weight 1 on A are called the *Legendre polynomials* for A. They have the explicit representation (not needed in what follows)

$$Q_k(x) = \frac{\sqrt{2k+1}}{k!}\frac{d^k}{dx^k}[x(1-x)]^k, \quad k = 0,\dots,n. \tag{8.7}$$

For $k = 1,\dots,n$, if P is a polynomial of degree $< k$, $(M_n(Q_k),P) = (Q_k, M_n(P)) = 0$, since $M_n(P) \in \mathcal{P}_{k-1}$. We see that the $M_n(Q_k)$ are orthogonal to $\mathcal{P}_{k-1}$ and so they must be multiples of Q_k: $M_n(Q_k) = \lambda_{n,k}Q_k$, $k = 1,\dots,n$. For $k = 0$, this is also true if we put $\lambda_{n,0} = 1$. In order to compute the eigenvalues $\lambda_{n,k}$, we compare the coefficients of the leading terms in $M_n(Q_k)$ and in $\lambda_{n,k}Q_k$. If $Q_k = \alpha_k x^k + \dots$, the first is the same as the coefficient of the leading term of $M_n(\alpha_k x^k)$, which from (8.3) is found to be $\alpha_k \frac{(n+1)!n!}{(n+k+1)!(n-k)!}$; the second is $\lambda_{n,k}\alpha_k$. □

The eigenvalues $\lambda_{n,k}$ are positive and $\lambda_{n,k} \to 1$ for $n \to \infty$ and each fixed k. This implies that the series summability method Λ, which assigns to a series $\sum u_k$ its Λ-sum $\lim_{n\to\infty}\sum_{k=0}^n \lambda_{n,k}u_k$ is regular, that is , it sums each convergent series to its ordinary sum. We prove that for each $f \in L_1(A)$, $M_n(f)$ is the n-th Λ-sum of the Legendre orthonormal expansion $f \sim \sum_0^\infty (f,Q_k)Q_k$:

$$M_n(f) = \sum_{k=0}^n \lambda_{n,k}(f,Q_k)Q_k. \tag{8.8}$$

Indeed, $M_n(f) \in \mathcal{P}_n$ and is therefore of the form $M_n(f) = \sum_{k=0}^n \alpha_k Q_k$. By orthonormality,

$$\alpha_k = (M_n(f),Q_k) = (f, M_n(Q_k)) = \lambda_{n,k}(f,Q_k),$$

which yields (8.8). From this formula, we derive

Theorem 8.4. *The operators M_n commute:*

$$M_nM_m = M_mM_n, \qquad m,n = 0,1\dots. \tag{8.9}$$

Proof. Let $n \ge m$. Then,

$$M_nM_m(f,x) = \sum_{k=0}^m \lambda_{m,k}(f,Q_k)M_n(Q_k) = \sum_{k=0}^m \lambda_{m,k}\lambda_{n,k}(f,Q_k)Q_k,$$

and

$$\begin{aligned} M_nM_m(f,x) &= \sum_{k=0}^m \lambda_{m,k}(M_n(f),Q_k)Q_k \\ &= \sum_{k=0}^m \lambda_{m.k}(f,M_n(Q_k))Q_k = \sum_{k=0}^m \lambda_{m,k}\lambda_{n,k}(f,Q_k)Q_k. \end{aligned}$$ □

We now give an application of (ordinary) Bernstein polynomials to the theory of weak Haar spaces. Weak Haar spaces $\mathcal{F}$ (see §12 of Chapter 3) can be approximated by ordinary Haar spaces (which are easier to handle) if they consist of continuous functions on an interval I. The approximation can be achieved by a sequence of operators on $C(I)$. The simplest and most elegant approach is by means of the Bernstein operators B_m, $m = 0,1,\dots$. We shall assume that $I = A = [0,1]$.

Let $\mathcal{F} \subset C(A)$ be an n-dimensional weak Haar space. The intersection of $\mathcal{F}$ with a closed ball $B \subset C(A)$ is compact. As a general fact, convergent sequences of linear operators on a Banach space X converge uniformly on compact sets. Thus in particular

$$B_n(g) \to g \quad \text{uniformly on bounded subsets of } \mathcal{F}. \tag{8.10}$$

Moreover, for each $h \in C(A)$,

$$B_m(h_m) \to h \qquad \text{if} \qquad h_m \to h. \tag{8.11}$$

Theorem 8.5. *If $\mathcal{F} \subset C(A)$ is a weak Haar space of dimension n, then for some m_0, all $\mathcal{F}_m := B_m(\mathcal{F})$, $m \ge m_0$ are Haar spaces of $C(0,1)$ of dimension n for the open interval $(0,1)$.*

Proof. For an element $B_m(g)$, $g \in \mathcal{F}$ of $\mathcal{F}_m$, for which $B_m(g) \ne 0$, we have by Theorem 3.3 (iv),

$$Z_{(0,1)}(B_m(g)) \le S_{(0,1)}(g) \le n-1.$$

It is also easy to prove that for sufficiently large m, the space $\mathcal{F}_m$ has dimension n. □

As an application of this, we shall derive a form of Chebyshev's Theorem 5.1 of Chapter 3, which is valid for weak Haar spaces $\mathcal{F}$. By Haar's Theorem 4.3 of Chapter 3, no unicity property of best approximation can be expected. However, the characterization of best approximants remains valid *for at least one of them.*

Theorem 8.6 (Jones and Karlovitz [1970]). *Let $\mathcal{F} \subset C(A)$, $A = [a, b]$, be a weak Haar space of dimension n. For each $f \in C(A) \setminus \mathcal{F}$, there is an element $g \in \mathcal{F}$ of best approximation to f with the property that $f - g$ has $n + 1$ oscillations of magnitude $d := E(f, \mathcal{F})$:*

$$\begin{gathered} f(x_j) - g(x_j) = \sigma(-1)^j d, \\ j = 1, \dots, n+1, \ \sigma = \pm 1, \ 0 \le x_1 < \cdots < x_{n+1} \le 1. \end{gathered} \tag{8.12}$$

Proof. We can assume that $0 < a < b < 1$. Functions $g \in \mathcal{F}$ we extend by $g(a)$ onto $[0, a]$ by $g(b)$ onto $[b, 1]$, and obtain a new weak Haar space $\mathcal{F}^*$. By Theorem 8.5, $\mathcal{F}_m := B_m(\mathcal{F}^*)$ is a Haar space on $[a, b]$, and $\operatorname{dist}(f, \mathcal{F}_m)[a, b] \to \operatorname{dist}(f, \mathcal{F}) = d > 0$. For large m, $d_m > 0$; then let $B_m(g_m)$ be the function of best approximation to f from $\mathcal{F}_m$ on $[a, b]$. For some $x_{i,m}$ it satisfies

$$\begin{gathered} f(x_{i,m}) - B_m(g_m, x_{i,m}) = \sigma_m(-1)^i d_m, \\ i = 1, \dots, n+1, \quad \sigma_m = \pm 1, \quad a \le x_{1,m} < \cdots < x_{n+1,m} \le b. \end{gathered} \tag{8.13}$$

Now the $B_m(g_m)$ are bounded on $[a, b]$. For any bounded subset $\mathcal{F}_0$, the $\omega(g, t)$ for $g \in \mathcal{F}_0$ have a common majorant $\omega(t)$. Because of $\|g_m - B_m(g_m)\| \le 3\omega(1/\sqrt{m})$, the g_m are also bounded. By passing to subsequences, we can achieve, for $m \to \infty$, $g_m \to g \in \mathcal{F}$ on $[a, b]$, $x_{i,m} \to x_i$, $i = 1, \dots, n+1$, and that $\sigma_m = \sigma$ is constant. For $m \to \infty$, (8.13) yields (8.12). □

§ 9. Problems

9.1. If f is bounded on $[0, 1]$ and has a jump discontinuity at x_0, $0 < x_0 < 1$, then $B_n(f, x_0) \to \frac{1}{2}[f(x_0-) + f(x_0+)]$.

9.2. Let C_0 be the subset of $C[0, 1]$ which consists of all functions f with $f(0) = f(1) = 0$. For $f \in C_0$, $\|B_n(f)\| \le (1 - 2^{-n+1})\|f\|$.

9.3. Let B_n^r be the r-th iteration of the operator B_n: $B_n^1 := B_n$, $B_n^r := B_n(B_n^{r-1})$. Then $B_n^r(f) \to B_1(f)$, $r \to \infty$, for all $f \in C$ (Kelisky and Rivlin [1967]). For deeper theorems in this direction see Karlin and Ziegler [1970].

9.4. For a sequence of integers $k_n \to \infty$, one has $B_n^{k_n}(f) \to B_1(f)$ if and only if $k_n/n \to \infty$.

9.5. The eigenvalues of B_n are 1, 1, $(1 - 1/n)$, $(1 - 1/n)(1 - 2/n), \dots, (1 - 1/n) \cdots (1/n)$.

9.6. The constant C of Theorem 3.2 can be taken equal to 2.

9.7. Prove that $\|f - K_n(f)\|_1 \le Cn^{-1/2} \int_0^1 \sqrt{X} |df(x)|$ (Hoeffding [1971]).

§ 10. Notes

10.1. The asymptotically best constant in the inequality $|f(x) - B_n(f, x)| \le C\omega(f, \sqrt{X}/n)$ has been found by Esseen [1960]: the maximum for $f \in C[0, 1]$, of

$$\limsup_{n \to \infty} \frac{f(x) - B_n(f, x)}{\omega(f, \sqrt{X}/n)}$$

is equal to $C_1 = 2 \sum_{k=0}^{\infty} (k+1)(\lambda(2k+2) - \lambda(2k)) = 1.045564\dots$, where $\lambda(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-t^2/2}\, dt$. See also Sikkema [1971].

10.2. Let U_n be a sequence of linear operators which map $C[0, 1]$ into $\mathcal{P}_n$, which preserve linear functions and convexity of all orders. Then (Berens and DeVore [1980])

$$U_n(t^2, x) \ge B_n(t^2, x) = x^2 + n^{-1}X, \qquad 0 < x < 1,$$

with strict inequality unless $U_n = B_n$.

10.3. Amelkovich [1966] was the first to give a really simple proof of the saturation Theorem 5.1. He used the positivity of the operator B_n and the asymptotic formula of Voronovskaya (3.1). Lorentz and Schumaker [1972] extended this to saturation theorems for arbitrary positive operators which possess an asymptotic theorem.

10.4. Similar to the Bernstein operators are the operators of Szász-Mirakiyan:

$$S_n(f, x) := \sum_{k=0}^{\infty} f\Big(\frac{k}{n}\Big) e^{-nx} \frac{(nx)^k}{k!}, \qquad \text{for } [0, \infty)$$

and of Meyer-König and Zeller

$$T_n(f, x) := \sum_{k=0}^{\infty} f\Big(\frac{k}{n+k}\Big) x^k (1-x)^{n+1}, \qquad \text{for } [0, 1].$$

Most theorems of this chapter are valid also for the operators S_n, T_n; see for example Becker and Nessel [1978], Totik [1983_1], [1983_2]. But the convergence conditions are different. We quote only one example. For the space C_B of bounded continuous functions with the uniform norm, Totik [1983_1] shows that $S_n(f) \to f$ for a given $f \in C_B$ if and only if $f(x + h\sqrt{x}) - f(x) \to 0$ uniformly as $h \to 0$ or equivalently if $f(x^2)$ is uniformly continuous.

10.5. Bernstein polynomials are important in computer aided geometric design for the numerical construction and display of curves and surfaces. For example, to obtain a curve in $\mathbb{R}^d$, one begins with "control points" $a_k \in \mathbb{R}^d$ and defines $P(t) := \sum_{k=0}^{n} a_k p_{n,k}(t)$. Then $P(t)$, $0 \le t \le 1$, parametrically describes a curve which is controlled by the selection of the a_k. From the many algorithms for the display of $P(t)$ (see for example Farin [1988]), we describe only one due to de Castlejau. If $0 \le t \le 1$, then we can compute $P(t)$ from the recursive formula for $a_k(m) := a_k(m, t)$ given by

$$a_k(m) = (1-t)a_{k-1}(m-1) + t a_k(m-1), \qquad k = m, \dots, n,$$

where initially $a_k(0) := a_k$, $k = 0, \dots, n$. Then $P(t) = a_n(n, t)$.

10.6. Another example for Theorem 8.1 of Chapter 9 are the commuting Durrmeyer polynomials $M_n(f)$ of (8.2). Their approximation behavior is identical with those of the $K_n(f)$. Among other things, Heilmann [1988] has found their saturation theorem, expressed by

$$(*) \qquad \|f - M_n(f)\|_p \le C\frac{1}{n}|f|_Y\,, \qquad 1 \le p \le \infty\,.$$

The space Y is defined as follows. A function f belongs to Y if f' is absolutely continuous on $(0,1)$ and if $|f|_Y := \|\{x(1-x)f'\}'\|_p < \infty$ (here, L_∞ is to be replaced by C). Thus, $(*)$ is the Jackson inequality (B) of §8, Chapter 9 for the operators M_n. The Bernstein inequality

$$(**) \qquad |M_n f|_Y \le 2n\|f\|_p\,. \qquad 1 \le p \le \infty$$

has been proved by Bavinck [1976].

Chapter 11. Approximation of Classes of Functions, Müntz Theorems

§ 1. Approximation by Fourier Sums

For an individual function f, it is usually hopeless to ask for an exact formula for its approximation error. Therefore, we shall turn our attention to *classes of functions*. For example, let K be a subset of a Banach space X of functions on $\mathbb{T}$. If $E_n(f)$ is the error of approximation of $f \in X$ by trigonometric polynomials, we define *the error of approximation of K by $\mathcal{T}_n$* by means of the formula

$$(1.1) \qquad E_n(K)_X := \sup_{f\in K} E_n(f)\,.$$

It is the "worst functions" of K that determine $E_n(K)$.

A similar quantity is the error of approximation of K by a sequence of operators U_n, which map X (or at least K) into $\mathcal{T}_n$,

$$(1.2) \qquad E(K;U_n)_X = \sup_{f\in K} \|f - U_n f\|\,.$$

The obvious relation between (1.1) and (1.2) is that $E_n(K) \le E(K,U_n)$. In this and the next sections we shall give examples of exact or of asymptotically exact computations of the two quantities. The theorems of Favard of §4 in Chapter 7 are results of this type.

Let $B^r_\infty := B^r$ consist of all functions f on $\mathbb{T}$ for which $|f^{(r)}(t)| \le 1$ a.e. Equivalently, $f \in B^r$ are r-th periodic integrals of all measurable functions h that satisfy

$$(1.3) \qquad |h(t)| \le 1 \text{ a.e.}\,, \qquad \int_0^{2\pi} h(t)\,dt = 0\,.$$

Favard found $E_n(B^r)_\infty$. But even before him, Kolmogorov determined the error of approximation $E(B^r, s_n)_\infty$ of the classes B^r, $r = 1, 2, \ldots$, in the uniform norm by means of the operator $s_n(f)$ of partial sums of Fourier series.

We have the representation $s_n(f,x) = \frac{1}{\pi}\int_0^{2\pi} f(x+t)D_n(t)\,dt$ with the Dirichlet kernel D_n. In addition to $D_{n,0} := D_n$ the following kernels will be used:

$$D_{n,1}(t) := \frac{t-\pi}{2} + \sum_{k=1}^{n} \frac{\sin kt}{k} = -\sum_{k=n+1}^{\infty} \frac{\sin kt}{k}, \quad 0 < t < 2\pi, \tag{1.4}$$

$$D_{n,2s}(t) := (-1)^{s+1} \sum_{k=n+1}^{\infty} \frac{\cos kt}{k^{2s}} \tag{1.5}$$

$$D_{n,2s+1}(t) := (-1)^{s+1} \sum_{k=n+1}^{\infty} \frac{\sin kt}{k^{2s+1}}. \tag{1.6}$$

The kernels are continuous on $\mathbb{T}$ except for $D_{n,1}$, for which $D_{n,1}(2\pi-) - D_{n,1}(0+) = \pi$; we have $\frac{d}{dt} D_{n,r} = D_{n,r-1}$, $r = 1, 2, \ldots$. Integrating by parts, we obtain

$$s_n(f,0) = f(0) - \frac{1}{\pi} \int_0^{2\pi} f'(t) D_{n,1}(t)\, dt = f(0) + \frac{(-1)^r}{\pi} \int_0^{2\pi} h(t) D_{n,r}(t)\, dt.$$

Together with a function f, B^r contains any of its translations $f(\cdot + a)$. Therefore

$$E(B^r; s_n)_\infty =: \sup_{f \in B^r} |s_n(f,0) - f(0)| = \sup_h \frac{1}{\pi} \left| \int_0^{2\pi} h D_{n,r}\, dt \right|,$$

where the supremum is taken over all h that satisfy (1.3). Obviously,

$$E(B^r; s_n) \le \frac{1}{\pi} \int_0^{2\pi} |D_{n,r}|\, dt. \tag{1.7}$$

If r is odd, then $D_{n,r}$ is an odd function, and $h(t) = \operatorname{sign} D_{n,r}(t)$ satisfies both conditions (1.3). In this case we have the exact formula

$$E(B^r; s_n) = \frac{1}{\pi} \int_0^{2\pi} \left| \sum_{n+1}^{\infty} \frac{\sin kx}{k^r} \right| dt. \tag{1.8}$$

Theorem 1.1 (Kolmogorov [1935]). *For $r = 1, 2, \ldots$*

$$E(B^r; s_n)_\infty = \frac{4}{\pi^2} \frac{\log n}{n^r} + \mathcal{O}(n^{-r}). \tag{1.9}$$

Proof. (a) First let r be even, $r = 2s$. We note that

$$D_k(t) - D_{k-1}(t) = \cos kt$$

$$D_k^*(t) - D_{k-1}^*(t) = D_k(t), \qquad D_k^*(t) := \frac{1}{2} \left(\frac{\sin \frac{k+1}{2} t}{\sin \frac{t}{2}} \right)^2.$$

Let $\Delta_k := k^{-r} - (k+1)^{-r}$, $\Delta_k^2 := \Delta_k - \Delta_{k+1}$, then $\Delta_k = \mathcal{O}(k^{-r-1})$, $\Delta_k^2 = \mathcal{O}(k^{-r-2})$. We apply to $D_{n,r}$ given by (1.5), twice the Abel transformation:

$$\begin{aligned}(-1)^{s+1} D_{n,r}(t) &= \sum_{n+1}^{\infty} \frac{\cos kt}{k^{2s}} = -\frac{1}{(n+1)^r} D_n(t) + \sum_{n+1}^{\infty} \Delta_k D_k(t) \\ &= -\frac{1}{(n+1)^r} D_n(t) - \Delta_{n+1} D_n^*(t) + \sum_{k=n+1}^{\infty} \Delta_k^2 D_k^*(t).\end{aligned}$$

For an arbitrary h satisfying (1.3) we write

$$\frac{1}{\pi} \left| \int_0^{2\pi} h D_{n,r}\, dt \right| = \frac{1}{\pi (n+1)^r} \left| \int_0^{2\pi} h D_n(t)\, dt \right| + R_n. \tag{1.10}$$

Since $(k+1)^{-1} D_k^*$ is the Fejèr kernel, $\int_0^{2\pi} D_k^*\, dt = \mathcal{O}(k)$, and the remainder satisfies $R_n = \mathcal{O}(n^{-r}) + \mathcal{O}(\sum k \cdot k^{-r-2}) = \mathcal{O}(n^{-r})$. An upper bound for the last integral in (1.10) is given by Theorem 2.1 of Chapter 9. In this way we get the inequality $\le$ in (1.9).

On the other hand, we select $h(t) = 0$ in $[0, \frac{2\pi}{2n+1}]$, $=$ sign $D_n(t)$ in $[\frac{2\pi}{2n+1}, 2\pi]$. This function satisfies (1.3), and the second integral in (1.10) becomes $\int_{2\pi/(2n+1)}^{2\pi} |D_n|\, dt$. This is $(4 \log n)/\pi + \mathcal{O}(1)$. In this way we obtain (1.9).

(b) For odd $r = 2s + 1$, the proof is slightly different. Here we use the functions

$$\Phi_k(t) := -\frac{\cos \frac{2k+1}{2} t}{2 \sin \frac{t}{2}}, \qquad \Psi_k(t) := -\frac{\sin(k+1)t}{4 \sin^2 \frac{t}{2}},$$

and have

$$\Phi_k(t) - \Phi_{k-1}(t) = \sin kt, \qquad \Psi_k(t) - \Psi_{k-1}(t) = \Phi_k(t).$$

We apply the Abel transformation twice:

$$(-1)^{s+1} D_{n,r}(t) = -\frac{1}{(n+1)^r} \Phi_n(t) - \Delta_{n+1} \Psi_n(t) + \sum_{k=n+1}^{\infty} \Delta_k^2 \Psi_k(t).$$

Now

$$\left(\int_0^{1/n} + \int_{2\pi - 1/n}^{2\pi} \right) |D_{n,r}|\, dt = \mathcal{O}(n^{-r}),$$

and we can write

$$\int_0^{2\pi} |D_{n,r}|\, dt = \mathcal{O}(n^{-r}) + \frac{1}{(n+1)^r} \int_{1/n}^{2\pi - 1/n} |\Phi_n(t)|\, dt + R_n. \tag{1.11}$$

Because of

$$\int_{1/n}^{2\pi - 1/n} |\Psi_n(t)|\, dt = \mathcal{O}(1) \int_{1/n}^{2\pi - 1/n} \frac{dt}{\sin^2 \frac{t}{2}} = \mathcal{O}(n),$$

the remainder is $R_n = \mathcal{O}(n^{-r})$. Now the second integral in (1.11) is again $\frac{4}{\pi} \log n + \mathcal{O}(1)$. The proof is completed as in (a). □

§ 2. Saturation Classes

For the partial sums $s_n(f)$ of the Fourier series of f we do not have $s_n(f) \to f$ in $C(\mathbb{T})$ for each f. But we do have fast convergence of $s_n(f)$ for smooth functions f. Actually, *the convergence* $\|s_n(f) - f\|_\infty \to 0$ *can be arbitrarily fast* for some f (that are not trigonometric polynomials): it is sufficient to take $f(t) = \sum(a_k \cos kt + b_k \sin kt)$, where a_k, b_k converge to zero sufficiently fast without being zero.

In contrast, some useful operators, such as the Fejèr sums $\sigma_n(f)$, and the Bernstein polynomials $B_n(f)$ converge to f for all continuous functions f. But the convergence is slow. Since these operators are positive, by Theorem 4.1 of Chapter 9 convergence cannot be as fast as $o(n^{-2})$. What is the best possible order of convergence, and for which f is it achieved? *Saturation theorems* answer this question. We shall treat these theorems for $C(\mathbb{T})$ here, for Bernstein operators they have been discussed in Chapter 10. We follow the definitions of Favard, reformulated by DeVore [A-1972]. We will discuss our problems only in the uniform norm.

Let $\phi_n > 0$, $n = 1, 2, \dots, \phi_n \to 0$. The sequence $\{U_n\}$ of uniformly bounded operators on $C(\mathbb{T})$ is *saturated of order* ϕ_n if

(i) $\liminf_{n\to\infty} \frac{\|f - U_n f\|}{\phi_n} = 0$ *implies that f is a constant.*
(ii) *There exists a non-constant $f_0 \in C(\mathbb{T})$ for which* $\|f_0 - U_n f_0\| = \mathcal{O}(\phi_n)$.

If this holds, then the set of all f for which $\|f - U_n f\| = \mathcal{O}(\phi_n)$, is called *the saturation class* $\mathfrak{S}(U_n)$ *of the sequence* U_n, which corresponds to the order (ϕ_n), and (ϕ_n) is called a *saturation sequence*. First saturation theorems have been given by Voronovskaya (Theorem 3.1 of Chapter 10) in 1932 and by Alexits [1941].

If (ϕ_n) is a saturation sequence and $\phi'_n \sim \phi_n$, then also (ϕ'_n) is one. The definition of the saturation sequence, which we have selected, ensures that if (ϕ_n) exists, it is *essentially unique. Namely, if (ϕ_n), (ϕ'_n) are two saturation sequences for the U_n, then* $\phi_n \sim \phi'_n$. For instance, we must have $\phi'_n \le \mathrm{Const}\phi_n$. Otherwise, for some $n_k \to \infty$, $\phi'_{n_k}/\phi_{n_k} \to \infty$, and we would have $\liminf (\|f_0 - U_n f_0\|/\phi'_n) = 0$ for the non-constant f_0 postulated by (ii) for the sequence (ϕ_n).

As candidates for saturation we consider sequences of convolution operators $U_n f = f * K_n$, where K_n is a continuous even function on $\mathbb{T}$ with mean value 1 and the Fourier series

$$K_n(t) := \frac{1}{2} + \sum_{k=1}^{\infty} \rho_n(k) \cos kt. \tag{2.1}$$

The operators $U_n(f)$ are given by

$$U_n(f,x) = \frac{1}{\pi}\int_{\mathbb{T}} f(t) K_n(x-t)\,dt = \frac{1}{\pi}\int_{\mathbb{T}} f(x+t) K_n(t)\,dt\,. \tag{2.2}$$

If a_k, b_k are the Fourier coefficients of f, those of $U_n(f)$ are $\rho_n(k)a_k$, $\rho_n(k)b_k$ and so the Fourier series of $U_n(f)$ is

$$U_n(f,x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} \rho_n(k)\,(a_k \cos kx + b_k \sin kx)\,. \tag{2.3}$$

Often the operators U_n are positive, this amounts to $K_n(t) \ge 0$, $t \in \mathbb{T}$.

For the functions $f_k(t) = \cos kt$, from (2.1) and (2.3) $f_k - U_n(f_k) = (1 - \rho_n(k))\,f_k$. Evaluating this at 0, we obtain, if $e_0(x) := 1$, $x \in \mathbb{R}$,

$$1 - \rho_n(k) = 1 - U_n(f_k, 0) = U_n(e_0 - f_k, 0) = \frac{2}{\pi}\int_{\mathbb{T}} \sin^2 \frac{kt}{2} K_n(t)dt\,. \tag{2.4}$$

In particular, if K_n is positive, we see that $\rho_n(k) < 1$, $n, k = 1, 2, \dots$; if K_n is not positive, we *shall assume that* $\rho_n(k) \ne 1$ for all n, k. We also assume $U_n(f) \to f$ for all $f \in C(\mathbb{T})$. This implies in particular $\rho_n(k) \to 1$ for $n \to \infty$, $k = 1, 2, \dots$.

Theorem 2.1. (i) *The sequence (U_n) is saturated if and only if for some* $m = 1, 2, \dots$,

$$\liminf_{n\to\infty} \frac{|1 - \rho_n(k)|}{|1 - \rho_n(m)|} = \lambda_k > 0\,, \quad k = 1, 2, \dots. \tag{2.5}$$

(ii) *If* (2.5) *is satisfied, then U_n is saturated with order* $\phi_n = |1 - \rho_n(m)|$.

Proof. Let (2.5) be satisfied, we prove (i) and (ii) of the definition of the saturation property. If

$$\liminf_{n\to\infty} \frac{\|f - U_n f\|}{|1 - \rho_n(m)|} = 0\,,$$

then, taking the k-th Fourier coefficient of $f - U_n f$ we see that $\liminf \left|\frac{a_k - \rho_n(k)a_k}{1 - \rho_n(m)}\right| = 0$, hence $a_k = 0$, and similarly $b_k = 0$, $k = 1, 2, \dots$. This shows that f is constant. But for $f_0 = \cos mt$,

$$f_0 - U_n f_0 = (1 - \rho_n(m)) \cos mt = \mathcal{O}(|1 - \rho_n(m)|)\,.$$

This yields (ii) of the definition, so that U_n is saturated with order $|1 - \rho_n(m)|$.

Let now U_n be saturated with order ϕ_n. Since $f(t) = \cos kt$ is not a constant, and $f - U_n(f) = (1 - \rho_n(k))\,f$, we derive from (i) of the definition that $|1 - \rho_n(k)|\,\phi_n^{-1} \ge C_k$ for some $C_k > 0$ for each $k = 1, 2, \dots$ and $n \ge n_k$. On the other hand, let f_0 be a non-constant function with the property $\|f_0 - U_n f_0\| \le C\phi_n$. The function f_0 has some non-zero Fourier coefficient, let $a_m \ne 0$. Then, taking Fourier coefficients,

$$|a_m|\,|1 - \rho_n(m)| \le C\phi_n$$

which implies (2.5). □

The following theorem partly describes the saturation class.

Theorem 2.2 (Sunouchi and Watari [1958-59]). *For some $m = 1, 2, \dots$ let*

$$\lim_{n\to\infty} \frac{1-\rho_n(k)}{1-\rho_n(m)} = \lambda_k \neq 0\,, \quad k = 1, 2, \dots\,, \tag{2.6}$$

and let $f \in \mathfrak{S}(U_n)$. If a_k, b_k are the Fourier coefficients of f, then

$$\sum_{k=1}^{\infty} \lambda_k \,(a_k \cos kt + b_k \sin kt) \tag{2.7}$$

is the Fourier series of some bounded function.

Proof. Let $f \in \mathfrak{S}(U_n)$. Then with $\rho_n := \rho_n(m)$,

$$\|f - U_n(f)\|_\infty = \mathcal{O}\left(|1-\rho_n|\right) .$$

It follows that all functions $g_n = (1-\rho_n)^{-1}\,(f - U_n(f))$ belong to a ball of a fixed radius in the space L_∞. Since the ball is weak*-compact, some sequence $g_{n_i} \to g$ converges weakly in L_∞. Consequently, $\int_{\mathbb{T}} g_{n_i} \cos kt\, dt \to \int_{\mathbb{T}} g \cos kt\, dt$ for $k = 1, 2, \dots$. For the coefficients α_k, β_k of g we have

$$\alpha_k = \lim_{i\to\infty} (1-\rho_{n_i})^{-1}\,(a_k - a_k \rho_{n_i}(k)) = a_k \lambda_k\,,$$

and similarly $\beta_k = b_k \lambda_k$. □

In the following theorem of Tureckii [1960] we have a complete description of the class $\mathfrak{S}(U_n)$ for certain sequences of positive operators. Let $\rho_n := \rho_n(1)$.

Theorem 2.3. *Let U_n be a sequence of positive convolution operators. If for each $k = 1, 2, \dots$*

$$\lim_{n\to\infty} \frac{1-\rho_n(k)}{1-\rho_n} = k^2\,, \quad k = 1, 2, \dots\,, \tag{2.8}$$

then U_n is saturated with order $1-\rho_n$, and the saturation class consists of all $f \in C(\mathbb{T})$ for which f' is absolutely continuous and $f' \in \mathrm{Lip}\,1$.

Proof. We need to prove only the last assumption. Let $f \in \mathfrak{S}(U_n)$. By Theorem 2.2,

$$\sum k^2\,(a_k \cos kt + b_k \sin kt)$$

is the Fourier series of a bounded function g. Then $-\sum(a_k \cos kt + b_k \sin kt)$ is the Fourier series of the second periodic integral of g. Hence f has the required properties.

Conversely, let $f' \in \mathrm{Lip}\,1$. From (4.15) of Chapter 9,

$$\|f - U_n f\| \le C\left(\|f\|\,(1-\rho_n) + \sqrt{1-\rho_n}\,\omega\left(f', \sqrt{1-\rho_n}\right)\right) \le \mathrm{Const}(1-\rho_n)\,.$$

□

As an example, let $J_{n,r}$, $r = 2, 3, \dots$, be the Jackson operator with the kernel $K_{n,r}$, given by (2.5) of Chapter 7. We use (2.4) and also Lemma 2.1 of Chapter 7 and see that $1-\rho_n(1) \sim n^{-2}$. Moreover, for each fixed $k = 1, 2, \dots$,

$$\begin{aligned} 1-\rho_n(k) &= \frac{1}{\pi}\int_{\mathbb{T}} 2\sin^2\frac{kt}{2} K_{n,r}(t)dt \\ &= \frac{2}{\pi}\int_{-\delta}^{\delta} \sin^2\frac{kt}{2} K_{n,r}(t)\,dt + \mathcal{O}(1)\int_{|t|\ge\delta} K_{n,r}(t)\,dt\,. \end{aligned}$$

The last integral is $\mathcal{O}(\lambda_{n,r}) = \mathcal{O}(n^{-3}) = o(1-\rho_n(1))$. For each $\varepsilon > 0$ and a sufficiently small δ, $\sin\frac{kt}{2}$ is between $k(1\pm\varepsilon)\sin\frac{t}{2}$ in $[-\delta, \delta]$. This yields (2.8). Therefore, *for all $r = 2, 3, \dots$ the sequence of the Jackson operators $J_{n,r}$ is saturated with order n^{-2}, and the saturation class consists of all f with $f' \in \mathrm{Lip}\,1$.*

Another example is provided by the de la Vallée-Poussin integral $V_n(f,x) := (f * K_n)(x)$ (see (8.12) of Chapter 9) with the kernel

$$K_n(t) = \frac{1}{2} + \sum_{k=1}^{n} \rho_n(k)\cos kt\,, \qquad \rho_n(k) = \frac{n!^2}{(n-k)!(n+k)!}\,.$$

One easily computes that (2.8) is valid. Therefore the saturation class of the family (V_n) consists of all f with $f' \in \mathrm{Lip}\,1$. The saturation order is $1-\rho_n(1) = (n+1)^{-1}$.

§ 3. Saturation of the Fejèr Operators

The operators σ_n do not satisfy (2.8). This situation requires a different approach. For the Fejèr kernel

$$F_n(t) = \frac{1}{2} + \sum_{k=1}^{n}\left(1-\frac{k}{n}\right)\cos kt\,,$$

we have $\rho_n(k) = 1 - \frac{k}{n}$, $1 \le k \le n$, $= 0$ for $k > n$. Since

$$(1-\rho_n(k))\,/\,(1-\rho_n(1)) \to k\,, \quad n \to \infty\,, \tag{3.1}$$

by Theorem 2.1, the operators $\sigma_n(f)$ are saturated with order $1-\rho_n(1) = n^{-1}$.

To determine the saturation class, we need the notion of the conjugate function. Let

$$\frac{a_0}{2} + \sum_{k=1}^{\infty}(a_k \cos kx + b_k \sin kx) =: \sum_{k=0}^{\infty} A_k(x)$$

be the Fourier series of an integrable function f. Then

$$\sum_{k=1}^{\infty}(b_k \cos kx - a_k \sin kx) =: \sum_{k=1}^{\infty} B_k(x) \tag{3.2}$$

is called the *conjugate series of* f. Let $\tilde{s}_n(f,x)$, $n = 1, 2, \dots$, denote the partial sums of (3.2), $\tilde{\sigma}_n(f,x) = (\tilde{s}_1 + \dots + \tilde{s}_{n-1})/n$, $n = 2, 3, \dots$, their arithmetic means. We wish to express $\tilde{s}_n$, $\tilde{\sigma}_n$ explicitly in terms of f.

The following formulas, similar to (1.11) and (1.12) of Chapter 1, are useful:

$$\tilde{D}_n(x) := \sin x + \cdots + \sin nx = \frac{\cos\frac{1}{2}x - \cos\left(n+\frac{1}{2}\right)x}{2\sin\frac{1}{2}x},$$

$$\tilde{F}_n(x) := \frac{\tilde{D}_1(x) + \cdots + \tilde{D}_{n-1}(x)}{n} = \frac{1}{2}\cot\frac{1}{2}x - \frac{\sin nx}{4n\sin^2\frac{1}{2}x}. \tag{3.3}$$

We leave their proof to the reader. Since

$$B_k(x) = \pi^{-1}\int_{-\pi}^{\pi} f(x+t)\sin kt\,dt,$$

we obtain

$$\tilde{s}_n(f,x) = \frac{1}{\pi}\int_0^{\pi}[f(x+t) - f(x-t)]\,\tilde{D}_n(t)dt,$$

$$\tilde{\sigma}_n(f,x) = \frac{1}{\pi}\int_0^{\pi}[f(x+t) - f(x-t)]\,\tilde{F}_n(t)dt. \tag{3.4}$$

If (3.2) is the Fourier series of some function $g \in L_1(\mathbb{T})$ then we call g the *conjugate function* of f and write $\tilde{f} := g$. It can be shown that the conjugate function $\tilde{f}$ exists and is in L_p whenever $f \in L_p$, $1 < p < \infty$ (M. Riesz); similarly if $f \in \operatorname{Lip}\alpha$, $0 < \alpha < 1$, then $\tilde{f} \in \operatorname{Lip}\alpha$ (Privalov). We will only need the following weaker version of the latter fact.

Lemma 3.1. *If $f \in \operatorname{Lip}\alpha$, $0 < \alpha \le 1$, then $\tilde{f}$ exists and is in $\operatorname{Lip}\beta$ for all $\beta < \alpha$; moreover we have the identity*

$$\tilde{f}(x) = \frac{1}{2\pi}\int_0^{\pi}[f(x+t) - f(x-t)]\cot\frac{1}{2}t\,dt. \tag{3.5}$$

Proof. It is clear that when $f \in \operatorname{Lip}\alpha$, the integral $I(x)$ in (3.5) exists. Also, if $\beta < \alpha$, then

$$|I(x+h) - I(x)| \le C\int_0^{\pi}\min(t^\alpha, h^\alpha)\,\frac{dt}{t} = \mathcal{O}(h^\beta),$$

so that I is in $\operatorname{Lip}\beta$. To verify (3.5) we calculate the Fourier coefficients $\tilde{a}_k$ and $\tilde{b}_k$ of I. Changing the order of integrations, we have

$$\begin{aligned}\tilde{a}_k &= \frac{1}{\pi}\int_{-\pi}^{\pi} I(x)\cos kx\,dx \\ &= \frac{1}{2\pi}\int_0^{\pi}\cot\frac{1}{2}t\,dt\frac{1}{\pi}\int_{-\pi}^{\pi} f(u)\,[\cos k(u-t) - \cos k(u+t)]\,du \\ &= b_k\frac{1}{\pi}\int_0^{\pi}\frac{\sin kt\cos\frac{1}{2}t\,dt}{\sin\frac{1}{2}t} \\ &= b_k\frac{1}{\pi}\int_0^{\pi}(D_k + D_{k-1})\,dt = b_k, \quad k = 1, 2, \ldots,\end{aligned}$$

and $\tilde{a}_0 = 0$; similarly $\tilde{b}_k = -a_k$, $k = 1, 2, \ldots$. □

After these preparations, we can formulate the following result.

Theorem 3.2 (Alexits [1941]). *The saturation class $\mathfrak{S}$ of the operators σ_n consists of all functions $f \in C(\mathbb{T})$ that have a conjugate function $\tilde{f} \in \operatorname{Lip}1$.*

Proof. First let $f \in \mathfrak{S}$, then $\|f - \sigma_n(f)\| \le Mn^{-1}$. By Bernstein's theorem (1.2) of Chapter 7, $f \in \operatorname{Lip}\alpha$ for each $0 < \alpha < 1$. Consequently $\tilde{f}$ exists. In view of (3.1) and Theorem 2.2,

$$-\sum_1^\infty k\,(a_k\cos kx + b_k\sin kx) \tag{3.6}$$

is the Fourier series of a function from L_∞. Integrating (3.6) termwise, we obtain (3.2). It follows that $\tilde{f}$ is an integral of a bounded function, hence it belongs to Lip 1.

Conversely, suppose that $\tilde{f}$ exists and that $|\tilde{f}(x+t) - \tilde{f}(x)| \le M|t|$. By Lemma 3.1, $\tilde{\tilde{f}}$ is continuous. Since $\tilde{\tilde{f}}$ and $-f + \frac{1}{2}a_0$ have the same Fourier coefficients, the two functions are identical. Moreover, $\tilde{\sigma}_n(\tilde{f}) = -\sigma_n(f) + \frac{1}{2}a_0$. Therefore we can interchange f and $\tilde{f}$ in (3.4) and (3.5):

$$f(x) - \sigma_n(f,x) = -\frac{1}{\pi}\int_0^{\pi}\left[\tilde{f}(x+t) - \tilde{f}(x-t)\right]\frac{\sin nt}{4n\sin^2(t/2)}dt = I_1 + I_2,$$

where the integrals I_1, I_2 are over the intervals $(0, \pi/n)$ and $(\pi/n, \pi)$, respectively. Since $|\sin nt| \le nt$,

$$|I_1| \le \frac{M}{2\pi}\int_0^{\pi/n}\frac{t^2}{\sin^2(t/2)}dt = \mathcal{O}(n^{-1}).$$

For I_2, we integrate by parts with the help of the function

$$\psi_n(t) := \int_t^{\pi}\frac{\sin nu}{\sin^2(u/2)}du:$$

$$\begin{aligned}|I_2| &\le \frac{1}{4\pi n}\left|\Big[(\tilde{f}(x+t) - \tilde{f}(x-t))\psi_n(t)\Big]_{t=\pi/n}^{t=\pi}\right| \\ &\quad + \frac{1}{4\pi n}\left|\int_{\pi/n}^{\pi}\psi_n(t)d_t\left\{\tilde{f}(x+t) - \tilde{f}(x-t)\right\}\right| \\ &\le \frac{M}{2n^2}\left|\psi_n\left(\frac{\pi}{n}\right)\right| + \frac{M}{2\pi n}\int_{\pi/n}^{\pi}|\psi_n(t)|\,dt.\end{aligned}$$

(The last estimate follows from the definition of the Stieltjes integral.) By the second mean-value theorem,

$$\int_a^b fg\,dt = g(a)\int_a^\xi f\,dt, \quad a \le \xi \le b,$$

if g is decreasing and positive and f is continuous. Therefore,

$$|\psi_n(t)| \le \frac{2}{n}\frac{1}{\sin^2(t/2)} \le \frac{2\pi^2}{nt^2}.$$

We obtain

$$|I_2| \le \frac{M}{n} + \frac{M\pi}{n^2}\int_{\pi/n}^{\pi}\frac{dt}{t^2} \le 2\frac{M}{n} = \mathcal{O}\left(\frac{1}{n}\right).$$

Hence

$$\|f - \sigma_n(f)\| = \mathcal{O}(n^{-1}). \qquad \square$$

§ 4. Theorems of Korneichuk

It is remarkable that the theorems of Favard of Chapter 7 can be extended to formulas which give the *exact error of approximation of the classes* $W^r H_\omega$, if ω is a concave modulus of continuity. They belong to Korneichuk (see his book [A-1976]). The following two theorems, also by Korneichuk, serve as an introduction.

Theorem 4.1. *For a concave ω one has for the trigonometric approximation, in the uniform norm on $\mathbb{T}$,*

$$E_{n-1}(H_\omega)_\infty = \frac{1}{2}\omega\left(\frac{\pi}{n}\right). \tag{4.1}$$

An extremal function f_ω with $E_n(f_\omega) = \frac{1}{2}\omega(\pi/n)$ is given by (4.3) *below.*

Proof. By Theorem 2.1 of Chapter 6, for $f \in H_\omega$, $K(f,h;C,\mathrm{Lip}\,1) = \frac{1}{2}\bar{\omega}(f,2h)$. There exists therefore a function $g \in \mathrm{Lip}\,1$ with

$$\|f-g\|_\infty + h\|g\|_{\mathrm{Lip}\,1} = \frac{1}{2}\bar{\omega}(f,2h) \le \frac{1}{2}\omega(2h). \tag{4.2}$$

By Favard's theorem of Chapter 7, there is a trigonometric polynomial T_{n-1} with $\|g - T_{n-1}\|_\infty \le \|g\|_{\mathrm{Lip}\,1}\pi/(2n)$. For these g and T_{n-1}, taking $h = \pi/(2n)$ in (4.2) we obtain

$$E_{n-1}(f) \le \|f - T_{n-1}\| \le \|f-g\| + \|g - T_{n-1}\| \le \frac{1}{2}\omega\left(\frac{\pi}{n}\right).$$

To prove the converse inequality, let $f := f_\omega$ be the odd function on $\mathbb{T}$ of period $2\pi/n$ given by

$$f_\omega(x) := \begin{cases} \frac{1}{2}\omega(2x), & 0 \le x \le \frac{\pi}{2n}, \\ \frac{1}{2}\omega(\frac{2\pi}{n} - 2x), & \frac{\pi}{2n} \le x \le \frac{\pi}{n}. \end{cases} \tag{4.3}$$

By (6.7) of Chapter 2, $f \in H_\omega$. This function has $2n$ points of maxima and minima on $\mathbb{T}$ with values $\pm\frac{1}{2}\omega(\pi/n)$ of alternating signs. By Chebyshev's theorem, the polynomial of degree $\le n-1$ of best approximation to f is zero. This shows that

$$E_{n-1}(H_\omega) \ge E_{n-1}(f_\omega) \ge \frac{1}{2}\omega\left(\frac{\pi}{n}\right). \qquad \square$$

We can also estimate the degree of approximation of the class H_ω when ω is not necessarily convex.

The following theorem implies in particular that the best possible constant in the simplest form of Jackson's inequality $E_n(f) \le C\omega(f, 1/n)$, for $p = \infty$ is $C = 1$:

Theorem 4.2. *For an arbitrary modulus of continuity ω on $\mathbb{T}$,*

$$E_{n-1}(H_\omega) \le \omega\left(\frac{\pi}{n}\right). \tag{4.4}$$

This cannot be improved: for each $M < 1$ and each sufficiently large $n \ge n_0(M)$, there is a modulus of continuity ω and a function $g \in H_\omega$ for which $E_{n-1}(g) > M\omega(\pi/n)$.

Proof. Inequality (4.4) follows from (4.1) and Lemma 6.1 of Chapter 2.

To prove the second part of the theorem, we show that for each $n = 1, 2, \dots$ and each $\varepsilon > 0$ there is a function $g \in C(\mathbb{T})$ with

$$E_{n-1}(g) \ge \left(\frac{2n-1}{2n} - \varepsilon\right)\omega\left(g, \frac{\pi}{n}\right). \tag{4.5}$$

The points $x_k = k\pi/n$, $k = 0, 1, \dots, n$, are π/n apart on $[0,\pi]$, but the points $y_k = k\pi/n - (n-k)\delta$, $k = 1, \dots, n$, are $(\pi/n) + \delta$ apart. We take a small $\delta > 0$, then $0 < y_1 - \delta < y_1 < \cdots < y_n = \pi$. Now

$$T_{n-1}(x) := \frac{1}{n}D_{n-1}(x) = \frac{1}{n}\frac{\sin(n-\frac{1}{2})x}{2\sin(x/2)}$$

is an even trigonometric polynomial, and

$$T_{n-1}(0) = \frac{2n-1}{2n}, \quad T_{n-1}(x_k) = \frac{(-1)^{k+1}}{2n}, \qquad k = 1, \dots, n. \tag{4.6}$$

If we take δ sufficiently small, we shall have

$$T_{n-1}(y_k) = \left(\frac{1}{2n} + \mu_k\right)(-1)^{k+1}, \quad |\mu_k| < \varepsilon, \qquad k = 1, \dots, n. \tag{4.7}$$

We construct an even continuous function g, defining it on $[0,\pi]$ in the following way: We put $g(y_k) = (-1)^{k+1}$, $k = 1, \dots n$, $g(y) = 0$ on each of the intervals $[0, y_1 - \delta]$, $[y_1 + \delta, y_2 - \delta], \dots, [y_{n-1} + \delta, y_n - \delta]$; and we make $g(y)$ linear on the remaining intervals. It is easy to understand that $\omega(g, \pi/n) = 1$. On the other hand, by (4.7) and the definition of g,

$$g(0) - T_{n-1}(0) = -\frac{2n-1}{2n},$$

$$g(y_k) - T_{n-1}(y_k) = \left(\frac{2n-1}{2n} - \mu_k\right)(-1)^{k+1}, \qquad k = 1, \dots, n.$$

Taking into account that g and T_{n-1} are even, we see that there are $2n$ points of $\mathbb{T}$ where $g(x) - T_{n-1}(x)$ takes values of alternating sign and of absolute value $\ge (2n-1)/(2n) - \varepsilon$. Theorem 5.2 of Chapter 3 gives $E_{n-1}(g) \ge (2n-1)/(2n) - \varepsilon$, and this is equivalent with (4.5). □

It is remarkable that Theorem 4.1 has a full extension to classes of differentiable functions. Without proof, we quote the following fundamental theorem of Korneichuk (see his books [A-1976], [A-1987]).

Theorem 4.3. *For each concave modulus of continuity ω and $r = 0, 1, 2, \dots$, one has on $\mathbb{T}$,*

$$E_{n-1}(W^r H_\omega)_\infty = n^{-r-1} \int_0^\pi \Phi_r(t)\omega'(t/n)\, dt, \tag{4.8}$$

where the Φ_r are polynomials on $[0, \pi]$ given by the recurrence relation

$$\Phi_r(x) = \frac{1}{2}\int_0^{\pi - x} \Phi_{r-1}(t)\, dt, \quad r = 1, 2, \dots, \quad \Phi_0(x) = \frac{1}{2}.$$

For example,

$$E_{n-1}(W^1 \operatorname{Lip}\alpha) = \frac{\pi^{1+\alpha}}{4(1+\alpha)} n^{-1-\alpha}, \quad E_{n-1}(W^2 \operatorname{Lip}\alpha) = \frac{\pi^{2+\alpha}}{8(2+\alpha)} n^{-2-\alpha}.$$

§ 5. Müntz' Theorem. Approximation of Monomials

Let $\Lambda : \lambda_0 < \lambda_1 < \dots < \lambda_n < \cdots$, $\lambda_n \to +\infty$ be a given sequence. We shall discuss the possibility of approximating all functions in $C[0, 1]$ or in $L_p[0, 1]$ by "Müntz polynomials"

$$P_n(x) = \sum_{k=0}^n a_k x^{\lambda_k}, \tag{5.1}$$

or, in other words, the problem of *completeness* of the sequence x^{λ_k} in these spaces. In general, we shall assume that $\lambda_0 \ge 0$, then all powers x^{λ_k} belong to these spaces. Sometimes we can assume less, for example $\lambda_0 > -\frac{1}{2}$ for L_2. If we want completeness in C, we must assume $\lambda_0 = 0$, for all other powers vanish at $x = 0$.

Let C_0 be the subspace of $C[0, 1]$ consisting of all functions that vanish at $x = 0$. Assume that X is a Banach function space embedded between C_0 and $L_1 : C_0 \subset X \subset L_1$, and let C_0 be dense in X. For example, X can be any L_p space, $1 \le p < \infty$. A sequence $\{x^{\lambda_k}\}$ which is complete in C_0, is also complete in all spaces X. Conversely, let it be complete in L_1, and let $f \in C_0$ be of the form $f(x) = \int_0^x g(t)\, dt$, $g \in L_1$. Approximating g in L_1 by a polynomial (5.1), we see that f is approximable in C by polynomials of the type $\sum_0^n b_k x^{\lambda_k + 1}$.

This partly explains the theorem of Müntz [1914]:

Theorem 5.1. *The functions x^{λ_k} are complete in any of the spaces X between C_0 and L_1 if and only if*

$$\sum_{k=1}^\infty \frac{1}{\lambda_k} = \infty; \tag{5.2}$$

for the space C one has to add the condition $\lambda_0 = 0$.

An elementary but powerful approach to the proof of this theorem is the approximation of fixed powers x^q by polynomials (5.1). Let $E_n(x^q)_X$ stand for the error of approximation of x^q by the polynomials $P_n(x)$. If we can prove that $E_n(x^q)_X \to 0$, then Weierstrass' theorem will reveal that all functions $f \in X$ are approximable by the P_n. Of course, the quantities $E_n(x^q)$ are interesting in themselves.

Proposition 5.2. *If (5.2) is satisfied, then $E_n(x^q)_\infty \to 0$ for all $q > 0$.*

The following simple *proof* was proposed by v. Golitschek. We assume that $q \ne \lambda_k$, $k = 0, 1, \dots$, and define the functions Q_n inductively:

$$Q_0(x) := x^q, \quad Q_n(x) := (\lambda_n - q)x^{\lambda_n} \int_x^1 Q_{n-1}(t) t^{-1-\lambda_n}\, dt, \quad n = 1, 2, \dots.$$

By induction, each Q_n is of the form $Q_n(x) = x^q - \sum_0^n a_{n,k} x^{\lambda_k}$. In the uniform norm, $\|Q_0\| = 1$, $\|Q_n\| \le |1 - \frac{q}{\lambda_n}| \cdot \|Q_{n-1}\|$, and because of (5.2),

$$E_n(x^q)_\infty \le \|Q_n\| \le \prod_{k=0}^n \left|1 - \frac{q}{\lambda_k}\right| \to 0, \qquad n \to \infty. \quad \square \tag{5.3}$$

The estimate (5.3) is not the best possible (see below), but it implies the sufficiency of the conditions of Theorem 5.1 for all the spaces involved.

It is remarkable that $E_n(x^q)_2$ can be computed exactly. In a Hilbert space H, let $f_1, \dots, f_n$ be linearly independent elements, and let X_n be the linear subspace which they span in H. We call the determinant

$$G(f_1, \dots, f_n) := \begin{vmatrix} (f_1, f_1), \dots, (f_1, f_n) \\ \cdots\cdots\cdots \\ (f_n, f_1), \dots, (f_n, f_n) \end{vmatrix} \tag{5.4}$$

the *Gram determinant* of the f_k.

Proposition 5.3. *The distance d from an element $g \in H \setminus X_n$ to X_n is given by*

$$d^2 = \frac{G(g, f_1, \dots, f_n)}{G(f_1, \dots, f_n)}. \tag{5.5}$$

Remark. Once (5.5) is established, it follows by induction that all Gram determinants (5.4) are > 0.

Proof. The best approximation $f \in X_n$ to g is defined by the orthogonality of $f - g$ to X_n. Now if $f = \sum_{i=1}^n a_i f_i$, then this orthogonality condition is equivalent to $f - g \perp f_k$, $k = 1, \dots, n$, that is, to

$$\sum_{i=1}^n a_i (f_i, f_k) = (g, f_k), \qquad k = 1, \dots, n. \tag{5.6}$$

Since this system of equations for the a_i must have a unique solution, its determinant $G(f_1, \dots, f_n)$ is not zero.

On the other hand, $d^2 = \|g - f\|^2$, and since $f \perp g - f$, $d^2 = (g - f, g - f) = (g, g) - (g, f)$.

To the system (5.6) we add the equation

$$\sum_{i=1}^n a_i (g, f_i) = (g, g) - d^2. \tag{5.7}$$

The column of free terms in the equations (5.6) and (5.7) is a linear combination of the columns of the coefficients of the a_i. Hence the determinant of order $n + 1$ formed by the columns, is zero. Expanding the determinant, we get (5.5). □

Theorem 5.4. *For q, $\lambda_k > -\frac{1}{2}$, $k = 0, 1, \dots$,*

$$E_n(x^q)_2 = \frac{1}{\sqrt{2q+1}} \prod_{k=0}^n \frac{|q - \lambda_k|}{q + \lambda_k + 1}. \tag{5.8}$$

Proof. First we note a formula of Cauchy,

$$\begin{vmatrix} \frac{1}{a_1+b_1}, \dots, \frac{1}{a_1+b_n} \\ \cdots\cdots\cdots \\ \frac{1}{a_n+b_1}, \dots, \frac{1}{a_n+b_n} \end{vmatrix} = \frac{\prod\limits_{i>k} (a_i - a_k)(b_i - b_k)}{\prod\limits_{i,k} (a_i + b_k)}. \tag{5.9}$$

To prove it, we subtract the last row of the determinant from each of the other rows; we can then factor out from the determinant the factor $\prod_{i=1}^{n-1}(a_n - a_i)/\prod_{k=1}^n (a_n + b_k)$. This leaves us with the determinant of the same matrix as in (5.9), except that in the last row the entries are now all 1. Next we subtract the last column from all the preceding ones, and extract the factors

$\prod_{i=1}^{n-1}(b_n - b_i)/\prod_{i=1}^n (a_i + b_n)$. Repeating this procedure n times we obtain the formula (5.9).

Now for the functions $f(x) = x^\lambda$, $g(x) = x^\mu$, λ, $\mu > -\frac{1}{2}$, we have

$$(f, g) = \int_0^1 x^\lambda x^\mu \, dx = \frac{1}{\lambda + \mu + 1}.$$

Using the identity (5.9), we see that the Gram determinant of the functions x^{λ_k}, $k = 0, \dots, n$ is equal to

$$G(x^{\lambda_0}, \dots, x^{\lambda_n}) = \frac{\prod\limits_{i>k} (\lambda_i - \lambda_k)^2}{\prod\limits_{i,k} (\lambda_i + \lambda_k + 1)},$$

with a similar formula for $G(x^q, x^{\lambda_0}, \dots, x^{\lambda_n})$. Substituting this into (5.5) and extracting the square root, we obtain (5.8). □

We can now prove that for L_2, condition (5.2) of Theorem 5.1 is necessary and sufficient for completeness. Indeed, if $\sum \lambda_k^{-1} < +\infty$, then each of the products $\prod(1 - q/\lambda_k)$ and $\prod(1 + (q+1)/\lambda_k)$ converges (to a value $\neq 0$), consequently $E_n(x^q) \not\to 0$ for any q which is not one of the λ_k. And if $\sum \lambda_k^{-1} = +\infty$, then the first product diverges to 0, the second diverges to $+\infty$, yielding $E_n(x^q) \to 0$ for all q.

The series $\sum \lambda_k^{-1}$ and $\sum (\lambda_k + 1)^{-1}$ converge or diverge at the same time. Therefore, these remarks and those preceding Theorem 5.1 give a proof of necessity, and a second proof of sufficiency in Theorem 5.1. There exist several complex-variable approaches to the problems of the Müntz polynomials (5.1), see for example Feinerman and Newman [A-1974].

Let $C[0, +\infty]$ be the space of continuous functions on $[0, +\infty]$, which have a finite limit for $t \to +\infty$. The transformation $x = e^{-t}$ maps the interval $[0, +\infty]$ onto $[0, 1]$. In this way, we obtain a variant of Theorem 5.1; exponential sums $\sum_{k=0}^n a_k e^{-\lambda_k t}$ approximate arbitrarily closely each function $f \in C[0, +\infty]$ if and only if $\lambda_0 = 0$ and $\sum_1^\infty \lambda_k^{-1} = +\infty$.

Finally, with the help of Theorem 5.4, we shall give another estimate of $E_n(x^q)$, also due to v. Golitschek [1970]:

Theorem 5.5. *For $q > 0$, $\lambda_k > 0$, $k = 1, 2, \dots$,*

$$E_n(x^q)_\infty \le \prod_{k=1}^n \frac{|q - \lambda_k|}{q + \lambda_k}. \tag{5.10}$$

Proof. For some $M > 0$, to be selected later, we put $\bar{q} = Mq$, $\mu_k = M\lambda_k$, $k = 1, 2, \dots$. For arbitrary coefficients c_k and some corresponding b_k we have, for each $0 \le x \le 1$,

$$\left|x^{\bar q+\frac12}-\sum_{k=1}^n b_k x^{\mu_k+\frac12}\right| = (\bar q+\frac12)\left|\int_0^x \left(t^{\bar q-\frac12}-\sum_1^n c_k t^{\mu_k-\frac12}\right)dt\right|$$
$$\le (\bar q+\frac12)\sqrt{x}\left\|t^{\bar q-\frac12}-\sum_1^n c_k t^{\mu_k-\frac12}\right\|_2 .$$

We now take the c_k which minimize the last norm and obtain, using (5.8),

$$\left|x^{Mq}-\sum_{k=1}^n b_k x^{M\lambda_k}\right| \le \frac{\bar q+\frac12}{\sqrt{2\bar q}}\prod_{k=1}^n \frac{|q-\lambda_k|}{q+\lambda_k} .$$

We take $M = 1/(2q)$, then the right side becomes the product in (5.10). On the left side, we put $x^M = u$ and take the supremum for all $0 \le u \le 1$. □

§ 6. Case When $\sum \lambda_k^{-1} < +\infty$. Selection of Best Powers

In Müntz' theorem 5.1, if

(6.1) $$\sum_1^\infty \lambda_k^{-1} < +\infty ,$$

the closed linear subspace Y of L_2 spanned by the x^{λ_k}, is not equal to the whole space. The two cases, $\sum \lambda_k^{-1} = +\infty$ and (6.1) are quite different. In the first case, one can omit from the set $\{x^{\lambda_k}\}$ any finite number of powers except 1, and they still will span the whole of L_2. In the second case, the powers are *topologically free*, that is, each of them has a positive distance to the subspace spanned by all others. Indeed, if we denote by $d(q, \Lambda)$ the distance in $L_2[0,1]$ from x^q to the span of the x^{λ_k}, with λ_k from $\Lambda = \{\lambda_0, \dots, \lambda_n, \dots\}$, $\lambda_n \to +\infty$, then formula (5.8) yields

(6.2) $$d(q, \Lambda) = \frac{1}{\sqrt{2q+1}} \prod_{k=0}^\infty \frac{|\lambda_k - q|}{\lambda_k + q + 1} .$$

If (6.1) is satisfied, and $q \notin \Lambda$, the product converges, so that $d(q, \Lambda) > 0$. In the same way, $d(\lambda_k, \Lambda_k) > 0$, $k = 0, 1, \dots$, if Λ_k is the sequence Λ with λ_k omitted.

We shall assume that the λ_k are integers. This not only makes the functions z^{λ_k} analytic, but also insures that the λ_k do not come close to each other, since then $|\lambda_k - \lambda_i| \ge 1$, $i \ne k$. As we shall see, in this case, the closed span Y of the x^{λ_k} contains only functions analytic in $|z| < 1$. The following theorem 6.2 of Clarkson and Erdös [1943], completed by Korevaar [1947], describes Y exactly.

We assume (6.1), and first derive a lemma. We need the following quantities:

(6.3) $$N_m := N_m(\Lambda) := \{\text{ number of } \lambda_k \in \Lambda \text{ with } \lambda_k \le m\} ;$$

(6.4) $$r_m := r_m(\Lambda) := \sum_{\lambda_k > m} \lambda_k^{-1} .$$

Condition (6.1) implies

(6.5) $$\frac1m N_m(\Lambda) \to 0 , \quad r_m(\Lambda) \to 0 \quad \text{for} \quad m \to \infty .$$

If we replace Λ by one of its subsequences, then N_m and r_m become smaller, so that (6.5) is true uniformly for all subsequences of Λ. In particular, we need the subsequences Λ_k mentioned above.

Lemma 6.1. *For each $0 < \varepsilon < 1$ there is a $q_0 = q_0(\varepsilon, \Lambda)$ so that if $q \ge 1$ is an integer not in Λ_k, then*

(6.6) $$d(q, \Lambda_k) \ge (1-\varepsilon)^q , \quad q \ge q_0 , \quad k = 0, 1, \dots .$$

Proof. We write

$$d(q, \Lambda_k) = \frac{1}{\sqrt{2q+1}} \Pi' \Pi'' ,$$

where Π' is the product with terms

(6.7) $$\frac{|q-\lambda_i|}{q+\lambda_i+1} = \left|1 - \frac{2q+1}{q+\lambda_i+1}\right|$$

for $\lambda_i > 3q+1$, $i \ne k$, and Π'' is the product of these terms for $\lambda_i \le 3q+1$, $i \ne k$.

For Π', $0 < (2q+1)/(q+\lambda_i+1) < \frac12$, and since $0 < x < \frac12$ implies $1 - x > e^{-2x}$,

$$\Pi' \ge \exp \sum \frac{-2(2q+1)}{q+\lambda_i+1} \ge \exp\left\{-2(2q+1) \sum_{\lambda_i > 3q+1} \frac{1}{\lambda_i}\right\}$$
$$\ge \exp\{-6q r_{3q+1}(\Lambda)\} \ge (1-\varepsilon)^q$$

for all large q.

For the product Π'', the denominators of the fraction $|\lambda_i - q|/(\lambda_i + q + 1)$ are all $\le 4q + 2 \le 6q$, and their number is either $N := N_{3q+1}(\Lambda)$ or $N - 1$. The numerators are positive integers, and each value can be repeated at most twice. Therefore the product of the numerator is at least $([(N-1)/2]!)^2$. Since $k! \ge (k/e)^k$, $k = 1, 2, \dots$, we find, for some $a > 0$,

$$\Pi'' \ge \left(a\frac{N}{q}\right)^N = \left[\left(a\frac{N}{q}\right)^{N/q}\right]^q .$$

Now $(ax)^x \to 1$ if $x \to 0$, hence it is $> 1 - \varepsilon$ for all small x. It follows that $\Pi'' > (1-\varepsilon)^q$, $q \ge q_0$. This gives $d(q, \Lambda_k) \ge \frac{1}{\sqrt{q+1}}(1-\varepsilon)^{2q}$. Since $\varepsilon > 0$ is arbitrary, (6.6) follows. □

Theorem 6.2. *If $0 \le \lambda_0 < \lambda_1 < \cdots$ is a sequence of integers which satisfies (6.1), then the closed span of the powers x^{λ_k} in $L_2[0,1]$ or in $C[0,1]$ is the set of those functions f in these spaces which possess an analytic extension onto $|z| < 1$, and have a power series expansion*

$$f(z) = \sum_{k=0}^{\infty} a_k z^{\lambda_k}, \qquad |z| < 1. \tag{6.8}$$

Proof. (a) First let f belong to the closed span Y of the x^{λ_k} in L_2. Then there exist polynomials $P_n(x) = \sum_0^n a_{n,k} x^{\lambda_k}$ for which $\|P_n - f\|_2 \to 0$ and for some $M \ge 0$, $\|P_n\|_2 \le M$. We first derive a uniform majorization of the $a_{n,k}$:

$$|a_{n,k}| \le M(1-\varepsilon)^{-\lambda_k} \quad \text{for} \quad \lambda_k \ge q_0(\varepsilon). \tag{6.9}$$

This follows from $P_n(x) = a_{n,k}(x^{\lambda_k} - Q(x))$ (if $a_{n,k} \ne 0$), where Q is a polynomial with powers $\lambda_i \in \Lambda_k$. Therefore $|a_{n,k}| \le M d(\lambda_k, \Lambda_k)^{-1}$.

Next we show that the $a_{n,k}$ converge:

$$\lim_{n\to\infty} a_{n,k} = a_k. \tag{6.10}$$

Indeed, for any n, m, $P_n(x) - P_m(x) = (a_{n,k} - a_{m,k})(x^{\lambda_k} - Q(x))$, and we get

$$|a_{n,k} - a_{m,k}| \le \|P_n - P_m\|_2 / d(\lambda_k, \Lambda_k) \to 0$$

for $n, m \to \infty$. The a_k also satisfy (6.9), hence the power series $\sum a_k z^{\lambda_k} := g(z)$ has the radius of convergence at least 1. It remains to show that $g(x) = f(x)$ for each fixed $0 \le x < 1$. Let $\delta > 0$ be arbitrary. Because of estimate (6.9), taking ε so small that $x(1-\varepsilon)^{-1} < 1$, we can find K so large that

$$\Big|\sum_{k>K} a_{n,k} x^{\lambda_k}\Big| < \delta, \qquad \Big|\sum_{k>K} a_k x^{\lambda_k}\Big| < \delta,$$

and then N so large that

$$\sum_{k\le K} |a_{n,k} - a_k| x^{\lambda_k} < \delta, \qquad n \ge N.$$

It follows that $|P_n(x) - g(x)| < 3\delta$, $n \ge N$, so that $P_n(x) \to g(x)$. A subsequence of the P_n converges a.e. to f, hence we have $g = f$ a.e.

Our statement about the span Y of the x^{λ_k} in C also follows from this.

(b) We have now to show that each function $f \in L_2$ or $f \in C$, which has the representation (6.8), is approximable by polynomials P_n. The map $U_\theta f = f_\theta$, $0 < \theta < 1$, $f_\theta(x) = f(\theta x)$ for $0 \le x \le 1$, is a bounded linear operator of each of the spaces into itself, and for all f, $U_\theta f \to f$ for $\theta \to 1$ (for L_2 this follows from the Banach-Steinhaus theorem; the U_θ are bounded and $U_\theta f \to f$ if f is continuous). On the other hand, $f_\theta(x) = \sum_{k=0}^{\infty} a_k \theta^k x^{\lambda_k}$ is uniformly approximable on $[0,1]$ by its partial sums, which are polynomials of type (6.1). □

If $f \in C[0,1]$ is of the type (6.8), one cannot infer that it is uniformly approximable on $[0,1]$ by the partial sums of the power series. The power series may in fact diverge at $z = 1$. The following corollary depends on a Tauberian theorem.

Corollary 6.3. *If for some $q > 1$,*

$$\frac{\lambda_{k+1}}{\lambda_k} \ge q, \qquad k = 0, 1, \dots \tag{6.11}$$

then each function $f \in C[0,1]$ in the closed span of the powers x^{λ_k} is approximable by the partial sums of its power series (6.8).

Proof. Let $c_{\lambda_k} := a_k$, $c_n := 0$ for other n. The continuity of $f(x) = \sum_{k=0}^{\infty} a_k x^{\lambda_k} = \sum_{n=0}^{\infty} c_n x^n$ at $x = 1$ means Abel summability of the last "gap series". By a deep gap Tauberian theorem of Hardy and Littlewood (see Hardy [B-1949], p. 173), this implies convergence of the series for $x = 1$. From this one derives the uniform convergence of $\sum a_k x^{\lambda_k}$ for $0 \le x \le 1$. □

We discuss a new problem. Let $1 \le k < n$, we would like to select k integers $\Lambda : \lambda_1 < \cdots < \lambda_k$, $0 \le \lambda_i < n$ so that the error of uniform approximation on $[0,1]$ of x^n by polynomials $\sum_1^k a_i x^{\lambda_i}$ be as small as possible. How are the powers x^{λ_i} to be selected? For the space L_2 the question can be readily answered by means of the formula (5.8): the λ_i are to be selected as close as possible to n. That this is true also in the space $C[0,1]$, has been shown by Borosh, Chui and Smith [1977]:

(6.12) *The best powers $\lambda_i < n$ for the $C[0,1]$ approximation are $\lambda_1 = n - k, \dots, \lambda_k = n - 1$.*

This has been generalized by Pinkus (in the paper Smith [1978]) in two directions. First, one can replace the powers x^k by continuous functions $u_0, \dots, u_n$ on $[0,1]$, which form a Descartes system (see §11 of Chapter 3). One then approximates u_n by polynomials $P_\Lambda := \sum_1^k a_i u_{\lambda_i}$, with given $\Lambda : \lambda_1 < \cdots < \lambda_k$, where $k < n$, $0 \le \lambda_i < n$. Secondly, the norm can be taken in a fairly arbitrary Banach function space X; we shall restrict ourselves to spaces $X = L_p[0,1]$. We define

$$E_\Lambda(f)_p := \inf_{P_\Lambda} \|f - P_\Lambda\|_p. \tag{6.13}$$

The following is a generalization of (6.12):

Theorem 6.4. *For each Descartes system $(u_k)_1^n$ on $[0,1]$ and $1 \le p \le \infty$ one has*

$$E_{\Lambda^*}(u_n)_p \le E_\Lambda(u_n)_p, \tag{6.14}$$

whenever Λ and $\Lambda^* : \lambda_1^* < \cdots < \lambda_k^*$ satisfy $\lambda_i \le \lambda_i^*$, $i = 1, \dots, k$; equality is possible only when $\Lambda = \Lambda^*$.

Proof. Let $\Lambda \neq \Lambda^*$. Let P_Λ be the polynomial of best approximation to u_n. From Theorem 10.8 of Chapter 3, P_Λ interpolates u_n at some k points, $0 < x_1 < \cdots < x_k < 1$. We take the unique polynomial P_{Λ^*} which solves the equations $u_n(x_j) - P_{\Lambda^*}(x_j) = 0$, $j = 1, \ldots, k$. By the comparison lemma (Theorem 11.2 of Chapter 3), $|u_n(x) - P_{\Lambda^*}(x)| < |u_n(x) - P_\Lambda(x)|$, $x \in [0,1]$, $x \neq x_j$. Hence $E_{\Lambda^*}(u_n)_p \le \|u_n - P_{\Lambda^*}\|_p < \|u_n - P_\Lambda\|_p = E_\Lambda(u_n)_p$. □

§ 7. Problems

7.1. Find the saturation order and class for the convolution integral of Rogosinski which has the kernel

$$K_n(t) = \frac{1}{2}\left\{D_n\left(t + \frac{\pi}{2n+1}\right) + D_n\left(t - \frac{\pi}{2n+1}\right)\right\},$$

where D_n is the Dirichlet kernel.

7.2. For the approximation of the class $\mathrm{Lip}_1\,1$ on $[-1,1]$ by algebraic polynomials, one has $E_n(f) \le c_n(n+1)^{-1}$, where $\lim_{n\to\infty} c_n = \frac{\pi}{2}$ (Nikolskii; see Bernstein [A-1954, vol. 2, p. 399]).

7.3. Let a_k, b_k be the Fourier coefficients of $f \in L_2(\mathbb{T})$. Using Parseval's identity for the function $f(\cdot + t) - f(\cdot)$ prove that

$$E_n(f)_2^2 \le \frac{1}{2}\omega(f,t)_2^2 + \pi \sum_{k \ge n} (a_k^2 + b_k^2)\cos kt .$$

7.4. Derive from this that

$$E_n(f)_2^2 \le \frac{n}{4}\int_0^{\pi/n} \omega(f,t)_2^2 \sin nt\, dt ,$$

in particular that $E_n(f)_2 \le (1/\sqrt{2})\omega(f, \pi/n)_2$, where the constant $1/\sqrt{2}$ cannot be improved (Chernykh, see Korneichuk [A-1976, p. 237]).

7.5. For $f \in W_2^r(\mathbb{T})$, prove that

$$E_n(f)_2^2 \le n^{-2r} E_n\left(f^{(r)}\right)_2^2 .$$

7.6. Let $f(t) = \sum_{k=0}^\infty a_k \cos kt$, where a_k are positive, increasing, and satisfy $a_k^2 \le a_{k+1}a_{k-1}$, $k \ge 1$. Then the error of uniform approximation of f can be estimated by

$$E_n(f) \le \sum_{k=n+1}^{\infty} a_k \le 4e\, E_n(f)$$

(Newman and Rivlin [1976]).

7.7. Let $N = N(n) > n$, we approximate x^N on $[0,1]$ by polynomials of degree n. Prove that $E_n(x^N)_\infty \to 0$ if and only if $N = o(n^2)$.

7.8. Let c_k, $k = 1, \ldots, n$ be distinct, not real numbers. Show that the functions $(x - c_k)^{-1}$ form a Haar system.

§ 8. Notes

8.1. Theorems of Favard belong to the subject of this chapter, but because of their extreme importance, we have preferred to treat them in Chapter 7. An important generalization of the Jackson-Bernstein-Favard theory stems from Bernstein.

Let E_σ, $\sigma > 0$ be the space of all entire functions g of exponential type $\le \sigma (\sigma > 0)$. They are characterized by the assumption that the inequality $|g(z)| \le A(\varepsilon)e^{(\sigma+\varepsilon)|z|}$, $z \in \mathbb{C}$ is satisfied for all $\varepsilon > 0$. Bernstein has emphasized the importance of the subspace B_σ consisting of $g \in E_\sigma$ that are bounded on $(-\infty, \infty)$. There exist analogues of theorems of Jackson, Bernstein, Favard for the approximation of functions $f \in L_p(-\infty, +\infty)$ or of bounded continuous functions f by $g \in E_\sigma$. For example, in the uniform norm on $(-\infty, +\infty)$, one has

$$\|g^{(r)}\| \le \sigma^r \|g\| , \qquad g \in B_\sigma \tag{8.1}$$

and with $\mathcal{E}_\sigma(f) := \lim_{\varepsilon\to 0} \operatorname{dist}(f, E_{\sigma-\varepsilon})$ and the Favard constant K_r, for the unit ball B_∞^r of W_∞^r,

$$\mathcal{E}_\sigma(B_\infty^r) := \sup_{f \in B_\infty^r} \mathcal{E}_\sigma(f) = K_r \sigma^{-r} . \tag{8.2}$$

These results, obtained by complex methods by Bernstein, Akhiezer and others, reduce to theorems about trigonometric approximation for 2π-periodic functions f. See the books of Ahiezer [A-1956], Timan [A-1963].

8.2. Saturation theorems have been the subject of an extensive literature. Several approaches have been developed. The method of infinitesimal generators applies when the operators U_n form a commutative semi-group, that is, when $U_nU_m = U_mU_n = U_{m+n}$. For this approach see the book of Butzer and Berens [A-1967]. The more elementary approach via Theorems 2.1-2.3 is treated in DeVore [A-1972]. Very similar saturation theorems are valid for L_p-approximation. See the book of Butzer and Nessel [A-1971], with many details.

8.3. In his books [A-1976] and [A-1987], Korneichuk has results that parallel Theorem 4.3 and describe, in similar terms, the error of trigonometric approximation $E_n(B_1^r)_1$ of the ball B_1^r in the space $W_1^r(\mathbb{T})$. Here, as in Theorem 4.3, the extremal functions are known. They are obtained from the function $f\omega$ of (4.3) by means of r integrations with mean value zero.

The proofs of Korneichuk's theorems are based on a certain type of rearrangements of functions, which were devised by him specifically for this purpose; they do not reduce in a trivial manner to the equimeasurable rearrangements of functions of §2 of Chapter 2.

8.4. Several authors (for instance Ligun [1985], Zhuk [1967]) have tried to determine or estimate the best possible constant $C_{r,p}$ in the inequality

$$E_n(f)_p \le C_{r,p}\omega(f^{(r)}, \pi/n)_p , \qquad f \in W_p^r(\mathbb{T}) . \tag{8.3}$$

For instance, they have $C_{r,p} \le K_r/(2n^r)$ if r is odd, and one has equality if $p = 1$ or $p = \infty$; see Korneichuk [A-1976, p. 247].

8.5. Clarkson and Erdös proved that condition (5.2) is necessary and sufficient for the completeness of the powers x^{λ_k} in $C[a,b]$ for any interval $[a,b]$, $0 < a < b$.

Newman and Rivlin [1976] gave a very exact estimate for the degree of the uniform approximation of x^N on $[0,1]$ by polynomials of degree $\le n$ for $n < N$. See problems 7 and 8. One can obtain the latter result also from (5.8) and (5.10).

Chapter 12. Spline Approximation

§ 1. Introduction

Approximation by splines presents us with problems familiar from polynomial approximation. The Leitmotiv is the direct and inverse theorems and sometimes the determination of the approximation spaces (see §7 of Chapter 7). However, now there are two quite different types of spline approximation; the one where the knots of the approximating spline are fixed in advance and the second where the knots are allowed to depend on the function f to be approximated.

The main direct theorem for splines with fixed knots, that is, for the approximation from the space $\mathcal{S}_r(T, A)$, $A := [0,1]$, $T = (t_j)^n$, is given by the estimate (7.11) of Chapter 7:

$$E(f, \mathcal{S}_r(T))_p \le \|f - Q_T f\|_p \le C\omega_r(f, \delta)_p\,. \tag{1.1}$$

Here Q_T is a quasi-interpolant corresponding to the extended knot sequence $(t_j)_{-r+1}^{r+n}$ and δ is the mesh size

$$\delta := \delta_T := \max_{0\le j\le n}(t_{j+1} - t_j) \tag{1.2}$$

with $t_0 := 0$, $t_{n+1} := 1$.

We discuss in §§2-3 fixed knot spline approximation on $A = [0,1]$ for two special sequences of knots: the splines with equally spaced knots

$$T_n := \left\{\frac{1}{n}, \dots, \frac{n-1}{n}\right\} \tag{1.3}$$

and with their subsequence, the dyadic splines with the knots

$$\Delta_n := T_{2^n} = \{2^{-n}, \dots, 1 - 2^{-n}\}\,. \tag{1.4}$$

The direct theorems are the same in both cases, and follow from (1.1). But the inverse theorems are different which is a reflection of the nesting of the dyadic knots. The splines with knots (1.3) have particularly strong saturation theorems, caused by the mixing properties of the knots for different n. To the dyadic splines, which play an important role in this chapter, we return later, in §§7 and 8.

A study of the general properties of free knot spline approximation, touched upon in §7 of Chapter 7, continues in §4, followed in §5 by some

estimates of the error $\sigma_{n,r}(f)_\infty$ in the uniform norm. Estimation of this error in L_p spaces leads naturally to smoothness spaces in L_γ with $\gamma < 1$. The reason for this is as follows. For polynomial and fixed knot spline approximation, to achieve an error $E_n(f)_p = \mathcal{O}(n^{-\alpha})$, one imposes on f smoothness conditions of the type $\omega_r(f,t)_p = \mathcal{O}(t^\alpha)$ with the same p. For free knot splines, these requirements are less demanding. Smoothness conditions on f are needed only in the much larger space L_γ, $\gamma := (\alpha + 1/p)^{-1}$. Since it can happen that $\gamma < 1$, even though $p > 1$, we develop in §6 properties of moduli of smoothness in the spaces L_γ, $0 < \gamma < 1$.

An essential property of the dyadic splines in §6 is Theorem 6.5, which asserts the equivalence of the semi-norm of a function f in the Besov space $B_q^\alpha(L_p)$, $\alpha > 0$, $0 < q \le \infty$, $r = [\alpha] + 1$,

$$|f|_{B_q^\alpha(L_p)} := \left(\int_0^\infty \left[t^{-\alpha}\omega_r(f,t)_p\right]^q \frac{dt}{t}\right)^{1/q} \qquad 0 < q < \infty \tag{1.5}$$

(with a proper supremum for $q = \infty$) to some norm of the sequence $(d_{n,r}(f)_p)$ of the approximation errors.

Further theorems of free knot spline approximation in L_p, $0 < p < \infty$ ($p \ne \infty$), can be found in §9. They include the Jackson and the Bernstein theorems for this approximation (see §4 of Chapter 7) by Petrushev, and an explicit characterization of some approximation spaces by DeVore and Popov.

We have often occasion to use the discrete Hardy inequalities from §3 of Chapter 2. For a sequence $\boldsymbol{a} = (a_k)_{k\in\mathbb{Z}}$, we define $\|\boldsymbol{a}\|_{\theta,q} = (\sum[2^{k\theta}|a_k|]^q)^{1/q}$ for $\theta, q > 0$. If for the sequences $a_k, b_k \ge 0$ one has for some $\mu > 0$, either one of the two inequalities:

$$\begin{aligned} &(i) \quad b_k \le C_0\Big(\sum_{j=k}^{\infty} a_j^\mu\Big)^{1/\mu} \\ &(ii) \quad b_k \le C_0 2^{-k\lambda}\Big(\sum_{j=-\infty}^{k} (2^{j\lambda}a_j)^\mu\Big)^{1/\mu}, \end{aligned} \tag{1.6}$$

then

$$\|\boldsymbol{b}\|_{\theta,q} \le CC_0\|\boldsymbol{a}\|_{\theta,q} \tag{1.7}$$

provided $0 < \theta$ in the case of (i) and $0 < \theta < \lambda$ in the case of (ii).

In many cases we will be able to determine the approximation spaces of spline approximation for errors of order $n^{-\alpha}$, $\alpha > 0$.

If $E_n(f)_X$ is a sequence of some (linear or non-linear) approximation errors of a function f in a space X, then $A_q^\alpha(X)$, $0 < q \le \infty$, $\alpha > 0$ is defined as follows (see §9 of Chapter 7). This is the space of all $f \in X$ for which $E_n(f)_X$ is of order $n^{-\alpha}$ in the sense that

$$\|((E_n(f))_X)_1^\infty\|_{\alpha,q} = \begin{cases} \left\{\sum_{n=1}^\infty (n^\alpha E_n(f)_X)^q \frac{1}{n}\right\}^1/q\,, & 0 < q < \infty \\ \sup\limits_{n\ge1}(n^\alpha E_n(f)_X)\,, & q = \infty \end{cases} \tag{1.8}$$

is finite.

§ 2. Splines with Equally Spaced Knots

In this section, we discuss the Schoenberg spaces $\mathcal{S}_r(T_n, A)$ on $A := [0,1]$ with the simple basic knots

$$(2.1) \qquad T_n := (t_j)_1^{n-1}, t_1 = \frac{1}{n}, \ldots, t_{n-1} = \frac{n-1}{n}, \qquad n = 2, 3, \ldots$$

and with the auxiliary knots $t_{-r+1} := \cdots := t_0 = 0$, $t_n := \cdots := t_{n+r-1} := 1$. If $n = 1$, we take $T_1 := \emptyset$ and $\mathcal{S}_r(T_1, A) = \mathcal{P}_{r-1}$. From (1.1), for the error of approximation $e_{n,r}(f)_p$, $f \in L_p(A)$, by the elements of $\mathcal{S}_r(T_n, A)$, we obtain the estimate

$$(2.2) \qquad e_n(f)_p := e_{n,r}(f)_p \le C_r \omega_r(f, 1/n)_p, \qquad n = 1, 2, \ldots .$$

It should be noted that the sets T_n do not satisfy the inclusion $T_n \subset T_m$, for $m > n$ and therefore the $e_n(f)_p$ do not necessarily decrease for increasing n.

The following converse to (2.2) is due to DeVore and Richards [1972] for $p = \infty$, and to Butler and Richards [1972] for $1 \le p < \infty$ (see also Scherer [1974]).

Theorem 2.1. *For $r = 1, 2, \ldots$, and $f \in L_p(A)$, $1 \le p \le \infty$, $0 < \delta \le 1$, we have with $n := [1/\delta]$*

$$(2.3) \qquad \omega_r(f, \delta)_p \le C_r \begin{cases} \left(\dfrac{1}{n} \displaystyle\sum_{m=n}^{2n} e_m(f)_p^p \right)^{1/p}, & 1 \le p < \infty \\ \displaystyle\max_{n \le m \le 2n} e_m(f)_\infty, & p = \infty . \end{cases}$$

The proof of Theorem 2.1 requires a "mixing" Lemma 2.3, which asserts that for each point x, for a large number of m's, x falls in the "middle" of an interval $[\frac{j}{m}, \frac{j+1}{m}]$. First we consider the case when $x = i/N$ is a rational number.

Lemma 2.2. *If N and i are positive integers, $1 \le i \le N-1$, then at least $N/16$ of the integers $m = N, \ldots, 2N$ satisfy*

$$(2.4) \qquad \operatorname{dist}\left(\frac{i}{N}, T_m\right) \ge \frac{1}{6N} .$$

Proof. If $i = 1$ or $N-1$, any $m \in [6N/5, 12N/7]$ satisfies (2.4). Therefore, we can take $1 < i < N-1$. We first show that whenever $1 \le k \le N/2$ and $ik \equiv \mu \bmod N$ with $N/4 \le \mu \le 3N/4$, then (2.4) holds for $m := N + k$. In fact, then $|ik + \lambda N| \ge N/4$ for all integers λ. Hence, for any integer j, we have

$$\left| \frac{i}{N} - \frac{j}{m} \right| = \left| \frac{ik + (i-j)N}{N(N+k)} \right| \ge \frac{N/4}{N(N+k)} \ge \frac{1}{6N} .$$

It remains only to show that for each i there are at least $N/16$ integers k with the above properties. If $1 < i \le N/2$, then clearly we can take any integer k which satisfies $(a + \frac{1}{4})N \le ki \le (a + \frac{3}{4})N$ for some integer a with $0 \le a \le \frac{i}{2} - 1$. There are $[i/2]$ such a. Since each interval of length l contains at least $[l]$ integers, for each a there are at least $[N/2i]$ such k, hence altogether $[i/2][N/(2i)] \ge (i/4)(N/4i) = N/16$ numbers k. When $i \ge N/2$, any k which works for $N - i$ also works for i. $\square$

Lemma 2.3. *For each $x \in A$, and $n = 1, 2, \ldots$,*

$$(2.5) \qquad \operatorname{dist}(x, T_m) \ge \frac{1}{16n}$$

for at least $n/64$ values of m with $n \le m \le 2n$.

Proof. Assume that this is not true for some $x \in I$. There are $\ge \frac{1}{64}n$ values of N for which $n \le N \le \frac{65}{64}n$. Therefore (2.5) fails for one of them, that is,

$$(2.6) \qquad \operatorname{dist}(x, T_N) < \frac{1}{16n} .$$

This means that $|x - \frac{i}{N}| < \frac{1}{16n}$ for some $0 < i < N$. Now Lemma 2.2 implies that (2.4) is true for at least $N/16$ values of m in the range $n \le m \le (2 + \frac{1}{32})n$ and therefore for at least $n/32$ values of m in the range $n \le m \le 2n$. For these m, we have $\operatorname{dist}(i/N, T_m) \ge \frac{1}{6N}$, hence

$$\operatorname{dist}(x, T_m) \ge \operatorname{dist}\left(\frac{i}{N}, T_m\right) - \frac{1}{16n} \ge \frac{1}{6N} - \frac{1}{16n} \ge \frac{1}{16n} . \qquad \square$$

Proof of Theorem 2.1. We fix $r = 1, 2, \ldots$, $\delta_1 := \frac{1}{16rn}$ and take $0 < h \le \delta_1$. Then $A_{rh} := [0, 1 - rh]$ is the interval where all the $\Delta_h^r(f, x)$ are defined. For $n := [1/\delta]$ and $n \le m \le 2n$, we define B_m to be the set of all $x \in A$ for which $(x, x + \frac{1}{16n})$ contains no knots from T_m. Whenever (2.5) holds, $x \in B_m$, and therefore the B_m cover A_{rh} at least $n/64$ times. We obtain

$$(2.7) \qquad \frac{1}{n} \sum_{m=n}^{2n} \chi_{B_m}(x) \ge \frac{1}{64}, \qquad x \in A_{rh} .$$

Now if $h \le \delta_1$, then $rh \le \frac{1}{16n}$. If $S_m \in \mathcal{S}_r(T_m, I)$ satisfies $\|f - S_m\|_p = e_m(f)_p$, then $\Delta_h^r(f, x) = \Delta_h^r(f - S_m, x)$, $x \in B_m$, because S_m is a polynomial of degree $< r$ on $[x, x + rh]$. When $p = \infty$, this gives

$$|\Delta_h^r(f, x)| \le 2^r e_m(f)_\infty \le 2^r \max_{n \le k \le 2n} e_k(f)_\infty , \qquad x \in B_m .$$

Consequently $\omega_r(f, \delta_1)_\infty$ also has this upper bound. Since $\delta \le 16r\delta_1$, using (7.8) of Chapter 2, we derive (2.3) for $p = \infty$. Similarly for $1 \le p < \infty$, from (2.7),

$$\frac{1}{64}\int_0^{1-rh}|\Delta_h^r(f,x)|^p\,dx \le \frac{1}{n}\sum_{m=n}^{2n}\int_{B_m}|\Delta_h^r(f,x)|^p\,dx$$
$$= \frac{1}{n}\sum_{m=n}^{2n}\int_{B_m}|\Delta_h^r(f-S_m,x)|^p\,dx \le \frac{2^{rp}}{n}\sum_{m=n}^{2n} e_m(f)_p^p\,.$$

Recalling the definition (7.2), Chapter 2, of $\omega_r(f,t)_p$, we obtain (2.3). □

Inequality (2.3) implies the important *saturation theorem* (Gaier [1970]): for a function $f \in L_p(A)$, we have $e_{n,r}(f)_p = o(n^{-r})$ if and only if $\omega_r(f,t)_p = o(t^r)$, or what is the same thing (see Proposition 7.1 of Chapter 2) if f is a polynomial of degree $< r$. We recall that there are no similar saturation phenomena for polynomial approximation. The companion inequalities (2.2) and (2.3) also characterize the saturation class; a function $f \in L_p(A)$ is approximated with order $\mathcal{O}(n^{-r})$ if and only if $\omega_r(f,t)_p = \mathcal{O}(t^r)$, that is, (see Theorem 9.3 of Chapter 2), if $f \in W_p^r$ when $1 < p \le \infty$ and $f^{(r-1)} \in BV$ when $p = 1$.

It is also possible to characterize the approximation spaces $A^\alpha(L_p)$ and more generally the space $A_q^\alpha(L_p)$ by means of the Besov spaces of §10 of Chapter 2. The space $A_q^\alpha(L_p)$ is defined by (1.8) with $X = L_p$, $E_n(f) = e_n(f)_p$.

Theorem 2.4. *Let $r = 1,2,\dots$. For the approximation by splines $S \in \mathcal{S}_r(T_n, A)$ with equally spaced knots, we have the equivalences:*

(i) *$e_n(f)_p = o(n^{-r})$, $n \to \infty$, if and only if $f \in \mathcal{P}_{r-1}$, $1 \le p \le \infty$;*
(ii) *$e_n(f)_p = \mathcal{O}(n^{-r})$, $n \to \infty$, if and only if $f \in W_p^r$, when $1 < p \le \infty$ and $f \in W^{r-1}(BV)$ when $p = 1$;*
(iii) *$e_n(f)_p = \mathcal{O}(n^{-\alpha})$, $n \to \infty$, if and only if $f \in \mathrm{Lip}^*(\alpha, L_p)$, $0 < \alpha < r$, $1 \le p \le \infty$;*
(iv) *$f \in A_q^\alpha(L_p)$ if and only if $f \in B_q^\alpha(L_p)$, $0 < \alpha < r$, $1 \le p \le \infty$, $q > 0$.*

Proof. The "if" statements are all consequences of (2.2), while the reverse implications follow from (2.3). And (iv) is proved by means of the discrete norm (10.5) of Chapter 2 for the Besov spaces. □

§ 3. Approximation by Dyadic Splines

For $A = [0,1]$, the dyadic splines correspond to the simple knots

(3.1) $$\Delta_n := (t_j)_1^{2^n-1}\,,\ t_j := t_{j,n} := j2^{-n}\,,\ j = 1,\dots,2^n-1\,.$$

We take the auxiliary knots $t_j := j2^{-n}$, $j \le 0$, $j \ge 2^n$. The dyadic knot sequence is nested: $\Delta_0 \subset \Delta_1 \subset \cdots$. We let

$$d_n(f)_p := d_{n,r}(f)_p := E\big(f, \mathcal{S}_r(\Delta_n, A)\big)_p$$

be the error of approximation of $f \in L_p(A)$, $p > 0$ (or of $f \in C(A)$ if $p = \infty$). Dyadic splines have been investigated by Ciesielski [1973], to whom the main results of this section belong. We shall discuss direct and inverse theorems, and a useful representation formula for the (semi-)norm of a function in a Besov space (Theorem 3.4). As in §2, the direct approximation theorem follows from (1.1):

(3.2) $$d_{n,r}(f)_p \le C_r\omega_r(f, 2^{-n})_p\,, \qquad 1 \le p \le \infty\,.$$

Since $d_{n,r}(f)_p = e_{2^n,r}(f)_p$, we can expect occasionally better approximation by dyadic splines than by splines of §2, and weaker inverse theorems. For example, Theorem 2.5 shows that $e_{n,r}(f)_p = \mathcal{O}(n^{-\alpha})$ implies $f \in \mathrm{Lip}^*(\alpha, L_p)$ for $\alpha < r$, while from the corresponding relation $d_{n,r}(f)_p = \mathcal{O}(2^{-n\alpha})$ we shall be able to draw this conclusion only if $\alpha < r - 1 + \frac{1}{p}$. This requirement cannot be essentially improved, for if we take $f_0 = (x-a)_+^{r-1}$, where a is a dyadic point, then for all large n, $f_0 \in \mathcal{S}_r(\Delta_n)$ and $d_{n,r}(f_0) = 0$. On the other hand, $\omega_r(f_0,t)_p \ge Ct^{r-1+1/p}$ and therefore the inverse theorem does not hold for any $\alpha > r - 1 + 1/p$.

According to §3, Chapter 5, the splines $S \in \mathcal{S}_r(\Delta_n)$ are linear combinations of the translated dilates

(3.3) $$N_{j,n}(x) := N(2^n x - j)\,, \qquad j = 0, \pm 1, \dots$$

of the single B-spline $N(x) := N(x; 0, 1, \dots, r)$. This leads to the unique representation of $S \in \mathcal{S}_r(\Delta_n)$:

(3.4) $$S = \sum_{j\in\Lambda_n} c_{j,n}N_{j,n}\,, \qquad \Lambda_n := \{-r+1,\dots,2^n-1\}\,.$$

We can use the properties of the coefficients $c_{j,n}$ from Chapter 5 to derive a Bernstein type estimate for the smoothness of S. We let $A_{rh} := [0, 1-rh]$.

Lemma 3.1. *Let $p > 0$, $r = 1,2,\dots$, and $\lambda := r - 1 + 1/p$. Then for $n = 1,2,\dots$ and each dyadic spline $S \in \mathcal{S}_r(\Delta_n)$, we have*

(3.5) $$\|\Delta_h^r(S,x)\|_p(A_{rh}) \le C2^{n\lambda}h^\lambda\|S\|_p\,, \qquad 0 < h \le 1$$

with C depending only on r (and on p when p is close to 0).

Proof. We fix h and estimate the differences of the B-spline $N_{j,n}$. Let Γ denote the points $x \in A$ at a distance $\le rh$ from the knots of $N_{j,n}$. Then $\Delta_h^r(N_{j,n},x) = 0$, $x \notin \Gamma$. Since $N_{j,n}$ is in W_∞^{r-1}, we have

(3.6) $$|\Delta_h^r(N_{j,n},x)| \le C(2^n h)^{r-1}\,, \qquad x \in \Gamma\,.$$

The measure of Γ does not exceed $2r(r+1)h$. Applying the L_p norm on A_{rh}, we obtain from (3.6)

(3.7) $$\|\Delta_h^r(N_{j,n},x)\|_p(A_{rh}) \le C2^{n\lambda}2^{-n/p}h^\lambda\,.$$

Now to estimate the smoothness of S, we find from (3.4),

$$|\Delta_h^r(S,x)| \le \sum_{j\in\Lambda_n} |c_{j,n}|\,|\Delta_h^r(N_{j,n},x)|\,.$$

For each $x \in A$, at most $r(r+1)$ terms of the sum are nonzero. Hence, if $0<p<\infty$,

$$|\Delta_h^r(S,x)|^p \le C^p \sum_{j\in\Lambda_n} |c_{j,n}|^p |\Delta_h^r(N_{j,n},x)|^p\,. \tag{3.8}$$

Integrating this over A_{rh} and using (3.7), we obtain

$$\|\Delta_h^r(S)\|_p(A_{rh}) \le C 2^{n\lambda} h^\lambda \left(2^{-n}\sum_{j\in\Lambda_n} |c_{j,n}|^p\right)^{1/p},$$

which in view of the left inequality of (4.4), Chapter 5, gives (3.5). The case $p=\infty$ is similar. □

In what follows, $\mathcal{S}_r(\Delta_0,A) := \mathcal{P}_{r-1}$ and $d_0(f)_p := E_r(f,A)_p$. An inverse estimate to (3.2) is given by the following:

Theorem 3.2 (Ciesielski [1973]). *For $1\le p\le\infty$, $n=1,2,\dots$, and $r=1,2,\dots$, we have for $\lambda := r-1+1/p$ and $f\in L_p(A)$,*

$$\omega_r(f,2^{-n})_p \le C_r 2^{-n\lambda}\sum_{k=0}^n 2^{k\lambda} d_k(f)_p\,. \tag{3.9}$$

Proof. We fix p,n and r and let U_k denote a best L_p approximation to f from $\mathcal{S}_r(\Delta_k,A)$, $k=0,1,\dots$. We further define $u_k := U_k - U_{k-1}$, $k=0,1,\dots$, with $U_{-1}:=0$. If $0<h\le 2^{-n}$ and $x\in A_{rh}$, we can write

$$\Delta_h^r(f,x) = \Delta_h^r(f-U_n,x) + \sum_{k=0}^n \Delta_h^r(u_k,x)\,.$$

The term $k=0$ is zero. Since $u_k\in\mathcal{S}_r(\Delta_k,A)$, it follows from Lemma 3.1 that

$$\begin{aligned}\|\Delta_h^r(f)\|_p(A_{rh}) &\le 2^r d_n(f)_p + \sum_{k=1}^n \|\Delta_h^r(u_k)\|_p(A_{rh})\\ &\le C\left[d_n(f)_p + h^\lambda \sum_{k=1}^n 2^{k\lambda}\|u_k\|_p\right].\end{aligned}$$

Taking a supremum over $0\le h\le 2^{-n}$ and using the inequality $\|u_k\|_p \le d_k(f)_p + d_{k-1}(f)_p$, we arrive at (3.9). □

As in the last section, the companion inequalities (3.2) and (3.9) characterize the functions in the approximation space $A_q^\alpha(L_p)$.

Theorem 3.3. *Let $r=1,2,\dots$, $1\le p\le\infty$, and $\lambda := r-1+1/p$. For approximation by the dyadic splines $\mathcal{S}_r(\Delta_n,A)$, for all $0<\alpha<\lambda$ and $0<q\le\infty$, we have $A_q^\alpha(L_p) = B_q^\alpha(L_p)$. In other words, $f\in B_q^\alpha(L_p)$ is equivalent to*

$$\begin{cases} \left(\sum_{n=0}^\infty [2^{n\alpha} d_n(f)_p]^q\right)^{1/q} < \infty, & 0<q<\infty,\\ \sup_{n\ge 0} 2^{n\alpha} d_n(f)_p < \infty, & q=\infty,\end{cases} \tag{3.10}$$

and (3.10) *is an equivalent (quasi)semi-norm for $B_q^\alpha(L_p)$.*

The *proof* parallels that of Theorem 2.4. The discrete Hardy inequalities (1.7), the discrete Besov space norms (10.5) of Chapter 2, and (3.9) with $\alpha<\lambda$ are used for the inverse statement. □

Let Q_n, $n=0,1,\dots$ be quasi-interpolant operators of §4, Chapter 5, which correspond to the sequence of the sets of knots Δ_n, let $Q_{-1}:=0$, and let $R_n = Q_n - Q_{n-1}$, $n=0,1,\dots$. For each $f\in L_p(A)$, $1\le p<\infty$, $Q_n(f)\to f$ in L_p, and we have the representation $f=\sum_{n=0}^\infty R_n(f)$. Since $R_n(f)\in\mathcal{S}_r(\Delta_n,A)$, we can write $R_n(f,x) = \sum_{j\in\Lambda_n} a_{j,n}(f)N_{j,n}(x)$ where the $a_{j,n}$ are linear functionals on $L_1(A)$. Thus, each $f\in L_p(A)$ has the series representation

$$f = \sum_{n=0}^\infty \sum_{j\in\Lambda_n} a_{j,n}(f) N(2^n x - j)\,. \tag{3.11}$$

Formula (3.11) is an example of a wavelet decomposition. Each f is represented as a series of translated dilates of a single function φ (in this case N). The coefficient functionals $a_{j,n}$ are also translated dilates of two functionals corresponding to the parity of the index j.

Another important property of the wavelet decomposition (3.11) is that the membership of f in a Besov space is easily determined from the coefficient sequence $(a_{j,n}(f))$. To derive this, we note first that the Q_n are uniformly bounded projections from $L_p(A)$ onto $\mathcal{S}_r(\Delta_n,A)$. Hence, we have $d_n(f)_p \le \|f-Q_n(f)\|_p \le C d_n(f)_p$, $n=0,1,\dots$. Therefore, in (3.10), we can replace $d_n(f)_p$ by $\|f-Q_n(f)\|_p$ and obtain equivalent expressions.

Similarly, we can replace $d_n(f)_p$ by $\varepsilon_n(f)_p := \|Q_n(f)-Q_{n-1}(f)\|_p$, $n\ge 0$. Indeed, writing $Q_n(f)-Q_{n-1}(f) = (f-Q_{n-1}(f)) - (f-Q_n(f))$, we see that $\varepsilon_n(f)_p \le d_{n-1}(f)_p + d_n(f)_p$. Here, $d_{-1}(f) := \|f\|_p$. Hence, from Theorem 3.3, we have,

$$\left(\sum_{n=0}^\infty [2^{n\alpha}\varepsilon_n(f)_p]^q\right)^{1/q} \le C\|f\|_{B_q^\alpha(L_p)}\,. \tag{3.12}$$

On the other hand, for each $f\in L_p$, we have $f = Q_n(f) + \sum_{k=n}^\infty [Q_{k+1}(f) - Q_k(f)]$ with convergence in L_p. Hence, for $n\ge -1$,

$$(3.13)\quad d_n(f)_p \le \|f - Q_n(f)\|_p \le \sum_{k=n}^{\infty} \|Q_{k+1}(f) - Q_k(f)\|_p \le \sum_{k=n}^{\infty} \varepsilon_{k+1}(f)_p\,.$$

Thus, if the left side of (3.12) is finite for some $0 < \alpha < \lambda$, then Theorem 3.3 and the discrete Hardy inequality (1.7) applied to (3.13) yield

$$(3.14)\qquad \|f\|_{B_q^\alpha(L_p)} \le C\left(\sum_{n=0}^{\infty}[2^{n\alpha}\varepsilon_n(f)_p]^q\right)^{1/q}.$$

If we use (4.4) of Chapter 5, we get

$$\varepsilon_n(f)_p := \|Q_n(f) - Q_{n-1}(f)\|_p \sim \left(\sum_{j\in\Lambda_n} 2^{-n}|a_{j,n}(f)|^p\right)^{1/p}, \qquad n = 0, 1, \dots.$$

We can therefore replace $d_n(f)_p$ in (3.10) also by the right-hand expression. This allows us to give a characterization (similar to the Parseval theorem) of functions f of the class $B_q^\alpha(L_p)$.

Theorem 3.4. *For $\alpha, q > 0$, $1 \le p \le \infty$, a function f belongs to $B_q^\alpha(L_p)$ if and only if, for the coefficients $a_{j,n}(f)$ in (3.11), the following is finite:*

$$(3.15)\qquad \begin{cases} \left(\sum_{n=0}^{\infty} 2^{\alpha n q}\left(\sum_{j\in\Lambda_n} 2^{-n}|a_{j,n}(f)|^p\right)^{q/p}\right)^{1/q}, & 0 < q < \infty, \\ \sup_{n\ge 0} 2^{n\alpha}\max_{j\in\Lambda_n}|a_{j,n}(f)|, & q = \infty. \end{cases}$$

and (3.15) is an equivalent (quasi)norm for $B_q^\alpha(L_p)$.

For example, a function f is in $\mathrm{Lip}^*(\alpha, L_p)$ if and only if f has the representation (3.11) with $a_{j,n}(f) = \mathcal{O}(2^{-n\alpha})$.

§ 4. Splines with Free Knots

Until now, we have considered approximation by spline functions on $A := [0,1]$ whose knots are prescribed in advance. The situation regarding the error of approximation changes dramatically if the knots of the spline are allowed to depend on the function to be approximated.

For given $r = 1, 2, \dots$ and $n = 1, 2, \dots$, we let $\Sigma_n := \Sigma_{n,r}$ be the space of splines of order r with $\le n$ arbitrary knots (see §7 of Chapter 7). Thus, Σ_n is the union of all of the Schoenberg spaces $\mathcal{S}_r(T, A)$, with $T : t_1 \le \cdots \le t_n$. For convenience, the auxiliary knots for T will always be the endpoints 0 and 1. Thus for $S \in \Sigma_n$, we can use the representation

$$(4.1)\qquad S(x) = \sum_{j\in\Lambda_n} c_{j,T}N_{j,T}(x)$$

of Chapter 5 where $\Lambda_n := \{-r+1, \dots, n\}$. We denote by

$$(4.2)\qquad \sigma_n(f)_p := \sigma_{n,r}(f)_p := E(f, \Sigma_{n,r})_p$$

the L_p error of approximation of a function $f \in L_p$ on A.

The existence of a spline of best approximation for $f \in L_p(A)$ does not follow from Theorem 1.1 of Chapter 3 because Σ_n is not a linear space. We have nevertheless (see Rice [A-1969]):

Theorem 4.1. *Each function $f \in L_p(A)$, $0 < p < \infty$, or $f \in C(A)$, $p = \infty$, has an element $S^* \in \Sigma_{n,r}$ of best approximation, for which*

$$(4.3)\qquad \|f - S^*\|_p = \sigma_n(f)_p\,.$$

Proof. For fixed n, r, and p, we choose a sequence $S_m \in \Sigma_n$, $m = 1, 2, \dots$, for which $\|f - S_m\|_p \to \sigma_n(f)_p$. Then, $S_m = \sum_{-r<j\le n} c_jN_{j,T}$ with $c_j := c_j(m)$, $T := T(m) = (t_j(m))_{-r+1}^{n+r}$. We can assume that $t_j(m) \to t_j^*$, $j = -r + 1, \dots, n+r$. It can happen that some of the t_j^* coincide. We let Λ^* be the set of all j for which $t_{j+r}^* \ne t_j^*$ and put $T^* := (t_j^*)_{j\in\Lambda^*}$.

Let U_ε be the ε neighborhood of the points t_j^*. For sufficiently large m, all of the $t_j(m)$ will be in U_ε. In particular, if $j \notin \Lambda^*$, all points $t_j(m), \dots, t_{j+r}(m)$ will be inside an interval about t_j^* of length ε, so that for $t \notin U_\varepsilon$, we will have $N_{j,T(m)}(t) = 0$. It follows that $S_m(t) = \sum_{j\in\Lambda^*} c_j(m)N_{j,T(m)}(t)$, $t \notin U_\varepsilon$, for all sufficiently large m. For $j \in \Lambda^*$,

$$\lambda_j(m) := \frac{1}{r}(t_{j+r}(m) - t_j(m)) \to \frac{1}{r}(t_{j+r}^* - t_j^*) > 0\,.$$

From (4.9) of Chapter 5,

$$(4.4)\qquad \left(\sum_{j\in\Lambda^*} \lambda_j(m)|c_j(m)|^p\right)^{1/p} \le C\|S_m\|_p\,,$$

so that the $c_j(m)$ are bounded and we may assume that they converge: $c_j(m) \to c_j^*$. By Lemma 2.1 of Chapter 5, $N_{j,T(m)} \to N_{j,T^*}$ uniformly on $A \setminus U_\varepsilon$. If $S^* := \sum_{j\in\Lambda^*} c_j^*N_{j,T^*}$, then $S^* \in \Sigma_n$,

$$\|f - S^*\|_p(A \setminus U_\varepsilon) = \lim_{m\to\infty}\|f - S_m\|_p(A\setminus U_\varepsilon) \le \sigma_n(f)_p\,,$$

and (4.3) follows by letting $\varepsilon \to 0$. □

Splines of best approximation are in general not unique. Consider, for example, the continuous piecewise linear function f on $[0,1]$ which has knots at j/m, $j = 0, \dots, m$ and takes alternately the values ± 1 at these points. In the uniform norm, f has an infinite number of best approximations from $\Sigma_{n,2}$, $n \le m-2$. Indeed, the error of approximation is one and any $S \in \Sigma_n$, $\|S\|_\infty \le 1$, which does not disagree in sign with f is a best approximation. Cases of uniqueness are discussed in the book of Nürnberger [A-1988, p. 194].

For example, functions in $f \in C(A)$ with a continuous derivative $f^{(r)}(x) > 0$ on A, do have unique best uniform approximants. For L_2 approximation, see Notes 10.5.

The splines S^* of best approximation to $f \in L_p(A)$ from $\Sigma_{n,r}$ do not necessarily possess any smoothness. However, we can show that splines with the highest smoothness are arbitrarily close to being best approximants. The following theorem improves on a result of Burchard [1974].

Theorem 4.2. *Let $0 < p \le \infty$ and $n, r \ge 2$. If $f \in L_p(A)$, then*

$$\inf_{S \in \Sigma_{n,r} \cap C^{r-2}} \|f - S\|_p = \sigma_n(f)_p \tag{4.5}$$

Proof. Let $S^* \in \Sigma_n$ be a best L_p approximation to f from Theorem 4.1. If $S^* \in \mathcal{S}_r(T, A)$, with $T : t_1, \dots, t_m$, $m \le n$, then $S^* = \sum_{-r<j\le m} c_j N_j$ where $N_j := N_{j,T}$ are the B-splines for T. Now let $U := (u_j)_{-r+1}^{m+r}$ be a knot sequence whose basic knots $u_1, \dots, u_m$ from $(0,1)$ are all simple. If $S_U := \sum_{-r<j\le m} c_j N_{j,U}$, then $S_U \in \Sigma_n \cap C^{(r-2)}(A)$. If $p < \infty$, then Lemma 2.1 of Chapter 5 shows that S_U converges to S^* in $L_p(A)$ as $U \to T$ and hence we have (4.5).

When $p = \infty$, the above argument shows that any $S \in \Sigma_n \cap C$ can be approximated arbitrarly well by splines from $\Sigma_n \cap C^{(r-2)}$. Thus, we need only show that for each $\varepsilon > 0$, there is an $S \in \Sigma_n \cap C$ with

$$\|f - S\|_\infty \le \sigma_n(f) + \varepsilon\,. \tag{4.6}$$

For this, we write $S^* = S_0 + S_1$ with $S_1 \in C$ and $S_0 := \sum_{j\in\Lambda} a_j (x - t_j)_+^0$ with $\Lambda := \{j : t_{j+r-1} = t_j\}$. If $j \in \Lambda$, we choose $\delta > 0$ so small that $t_j + \delta < t_{j+r}$ for all $j \in \Lambda$ and let M_j be the B-spline of order $r-1$ with knots $t_j, \dots, t_j, t_j + \delta$ and let $\phi_j(x) := \int_0^x M_j(t)\,dt$. Then ϕ_j is a spline of order r with a knot of multiplicity $r-1$ at t_j and a simple knot at $t_j + \delta$. We define $S_\delta := \sum_{j\in\Lambda} a_j \phi_j$ and $\tilde{S}_\delta := S_\delta + S_1$. Then $\tilde{S}_\delta \in \mathcal{S}_r(\tilde{T})$ where $\tilde{T}$ is obtained from T by replacing t_{j+r-1} by $t_j + \delta$, $j \in \Lambda$. Thus, $\tilde{S}_\delta$ is in $\Sigma_n \cap C$ bcause with each new knot $t_j + \delta$, we have correspondingly reduced the multiplicity of t_j from r to $r-1$.

Now, the step function S_0 approximates the continuous function $f - S_1$ with error $\sigma_n(f)_\infty$. Also, S_δ agrees with S_0 except for the intervals $[t_j, t_j + \delta]$ on which S_δ takes values between $S_0(t_j-)$ and $S(t_j+)$. Hence, on $[t_j, t_j + \delta]$, S_δ approximates $f - S_1$ at least as well as one of the two numbers $S_0(t_j-)$ or $S_0(t_j+)$. We obtain

$$\|f - \tilde{S}_\delta\|_\infty \le \|f - S^*\|_\infty + \omega(f - S_1, \delta)$$

and (4.6) holds provided δ is sufficiently small. $\square$

We give next some elementary estimates for the error of free knot spline approximation. These will already show distinct improvement over linear methods. This comes from the possibility of placing the knots of the approximating spline where they are most needed, where f is less smooth. A simple result of this type is the following theorem of Kahane [1961] for approximation by splines in $\Sigma_n := \Sigma_{n,1}$. This means that we approximate f by step functions which assume at most $n+1$ different values.

For a closed subinterval I of $A := [0,1]$, and $f \in C(I)$, let $M_I := \max_I f$ and $m_I := \min_I f$. We begin with the following simple remark. If a step function S assumes at most k values on I, then

$$M_I - m_I \le 2k\|f - S\|_\infty(I)\,. \tag{4.7}$$

For the proof, we select points, $\xi_i \in I$ with $f(\xi_i) = m_I + i(M_I - m_I)/k$, $i = 0, \dots, k$. If we would have $|f(\xi_i) - S(\xi_i)| < \frac{(M_I - m_I)}{2k}$, for all i, then for all i, $S(\xi_i)$ and $S(\xi_{i+1})$ would be different, a contradiction.

Theorem 4.3 (Kahane). *For $f \in C(A)$ and $M > 0$, the following are equivalent:*

$$\begin{aligned} &\text{(i)} \quad \mathrm{Var}_A(f) \le M\,, \\ &\text{(ii)} \quad \sigma_n(f)_\infty \le \frac{M}{2n+2}\,, \qquad n = 1, 2, \dots\,. \end{aligned} \tag{4.8}$$

Proof. If (i) holds, then the continuity and subadditivity of $\mathrm{Var}_{[0,x]} f$ imply the existence of intervals $I_j := [t_j, t_{j+1}]$, $j = 0, \dots, n$, $0 =: t_0 < t_1 < \dots < t_{n+1} := 1$ for which $\mathrm{Var}_{I_j} f \le M/(n+1)$. The function $S_n(x) := (M_{I_j} + m_{I_j})/2$, $t_j < x < t_{j+1}$, $j = 0, \dots, n$, belongs to Σ_n and we have (ii) because

$$\|f - S_n\|_\infty(I_j) \le \frac{1}{2}(M_{I_j} - m_{I_j}) \le \frac{1}{2}\mathrm{Var}_{I_j} f \le M/(2n+2)\,.$$

Conversely, if (ii) holds there exist $S_n \in \Sigma_n$, $n = 1, 2, \dots$, with the property

$$\|f - S_n\|_\infty(A) \le \frac{M}{2n+2}\,.$$

Let $0 = x_0 < x_1 < \dots < x_m = 1$ be an arbitrary subdivision of A. If S_n takes $k_{j,n}$ values on $[x_{j-1}, x_j]$, then by (4.5)

$$\sum_{j=1}^m |f(x_j) - f(x_{j-1})| \le \sum_{j=1}^m 2k_{j,n}\frac{M}{2(n+1)} = \frac{M}{n+1}\sum_{j=1}^m k_{j,n}\,. \tag{4.9}$$

A value of S_n at an x_i could contribute to two of the $k_{j,n}$. Therefore, $\sum_{j=1}^m k_{j,n} \le n + m$ and (4.9) becomes

$$\sum_{j=1}^m |f(x_j) - f(x_{j-1})| \le M\left(1 + \frac{m-1}{n+1}\right)\,.$$

The sum here can be made arbitrarily close to $\mathrm{Var}_A f$. Letting $n \to \infty$, we obtain (i). $\square$

The proof of the inverse direction of Theorem 4.2 uses only that (ii) holds for sufficiently large n, therefore, as a corollary, we obtain a saturation statement:

Corollary 4.4. *For approximation by piecewise constants with n free knots, a function $f \in C(A)$ satisfies $\sigma_n(f)_\infty = o(n^{-1})$ if and only if f is constant.*

The implication (i) $\Rightarrow$ (ii) of Theorem 4.3 can be extended to higher derivatives by using the following simple remarks. Let S be a best $L_1(A)$ approximation to f' from $\Sigma_{n,r}$: $\|f' - S\|_1(A) = \sigma_{n,r}(f')_1$. We choose $n+1$ points $0 =: x_0 < x_1 < \cdots < x_n := 1$ such that $\|f' - S\|_1[x_{j-1}, x_j] = \frac{1}{n}\sigma_{n,r}(f')_1$ and define $\tilde{S}(x) := f(x_j) + \int_{x_j}^x S(t)\,dt$, for $x_j < x < x_{j+1}$. Then,

$$|f(x) - \tilde{S}(x)| \le \int_{x_{j-1}}^{x_j} |f'(t) - S(t)|\,dt = \frac{1}{n}\sigma_{n,r}(f')_1\,, \qquad x_{j-1} < x < x_j\,.$$

In addition to the knots of S, the spline $\tilde{S}$ has knots of multiplicity at most $r+1$ at each of the x_i. Hence $\tilde{S} \in \Sigma_{(r+2)n,r+1}$ and we obtain

$$\sigma_{(r+2)n,r+1}(f)_\infty \le \frac{1}{n}\sigma_{n,r}(f')_1\,. \tag{4.10}$$

Clearly, we can replace the L_∞ norm on the left side of (4.10) by any L_q norm, $1 \le q \le \infty$, and we can replace the L_1 norm on the right side of (4.10) by any L_p norm, $1 \le p \le \infty$.

As an application, we have the following result proved by Freud and Popov [1969] and independently by Subbotin and Chernykh [1970].

Theorem 4.5. *If $r = 1, 2, \ldots$ and if $f^{(r-1)}$ is of bounded variation on A, then*

$$\sigma_{n,r}(f)_\infty \le \frac{C_r \operatorname{Var}_A f^{(r-1)}}{n^r}\,. \tag{4.11}$$

Proof. This follows from (4.8)(ii) and repeated application of (4.10) and the remarks after (4.10). □

The estimates (4.6) and (4.11) already point out some of the essential features of nonlinear approximation. For linear approximation in $L_p(A)$ by splines with n equally spaced knots, one has the upper bound $Cn^{-r}|f|_{W_p^r}$ (Theorem 2.4). Thus the error of approximation is measured in the same norm as the smoothness. On the other hand, in the estimate (4.11) the approximation takes place in the L_∞ norm but the smoothness is measured in the L_1 sense (because $\operatorname{Var} f^{(r-1)}$ is closely related to $\|f^{(r)}\|_1$). The following theorem, given by Burchard [1974] for the case $p = \infty$, is another example of this type.

Theorem 4.6. *Let $r = 1, 2, \ldots$. If f is in $L_p(A)$, $1 \le p < \infty$ ($f \in C(A)$, $p = \infty$), $A := [0,1]$, and if $f \in C^r(0,1)$, and if $|f^{(r)}|$ has a monotone majorant $\varphi \in L_\gamma(A)$, $\gamma := (r + 1/p)^{-1}$ on A, then*

$$\sigma_{nr,r}(f)_p \le \frac{1}{r!\,n^r}\|\varphi\|_\gamma \qquad n = 1, 2, \ldots\,. \tag{4.12}$$

Proof. We follow de Boor [1973]. By considering $f(1-x)$ in place of $f(x)$, if necessary, we may assume that φ is decreasing on $(0,1)$. Furthermore, we can assume that $p < \infty$ (the case $p = \infty$ will then follow by passing to the limit $p \to \infty$). We shall use the following inequality from Lorentz [A-1953, p. 66]:

$$\int_0^x g(t)t^{\lambda-1}\,dt \le \frac{1}{\lambda}\left(\int_0^x g(t)^{1/\lambda}\,dt\right)^\lambda, \qquad 0 \le x \le c\,, \tag{4.13}$$

where $g(t) \ge 0$ is decreasing on some interval $[0, c]$ and $\lambda \ge 1$. In fact, the derivative of the right side of (4.13) is larger than $g(x)x^{\lambda-1}$ which is the derivative of the left side.

We let $M := \|\varphi\|_\gamma$ and select points $0 =: x_0 < x_1 < \cdots < x_n := 1$ so that for the intervals $I_j := [x_{j-1}, x_j]$, $j = 1, \ldots, n$,

$$\int_{I_j} \varphi(x)^\gamma\,dx = \frac{1}{n}\int_0^1 \varphi(x)^\gamma\,dx = \frac{M^\gamma}{n}\,. \tag{4.14}$$

Let P_j be the Taylor polynomial for f of degree $r-1$ at the point x_j. Then, for $x \in I_j$ (see (5.6) of Chapter 2):

$$(f - P_j)(x) = \frac{(-1)^r}{(r-1)!}\int_{I_j} (t-x)_+^{r-1} f^{(r)}(t)\,dt\,.$$

Now, for each $t \in I_j$, the $L_p(I_j)$ norm of $(t-x)_+^{r-1}$ with respect to x does not exceed $(t - x_{j-1})^{r-1+1/p}$. By means of Minkowski's inequality and (4.13), we derive

$$\begin{aligned} \|f - P_j\|_p(I_j) &\le \frac{1}{(r-1)!}\int_{I_j} |\varphi(t)|\ \|(t-x)_+^{r-1}\|_p\,dt \\ &\le \frac{1}{(r-1)!}\int_{I_j} \varphi(t)(t - x_{j-1})^{r-1+1/p}\,dt \\ &\le \frac{1}{r!}\|\varphi\|_\gamma(I_j)\,. \end{aligned} \tag{4.15}$$

The spline function $S(x) := P_j(x)$, $x_{j-1} < x < x_j$ is in $\Sigma_{nr,r}$ and satisfies

$$\begin{aligned} \|f - S\|_p &\le \left(\sum_{j=1}^n \|f - P_j\|_p^p(I_j)\right)^{1/p} \le \frac{1}{r!}\left(\sum_{j=1}^n \|\varphi\|_\gamma^p(I_j)\right)^{1/p} \\ &= \frac{Mn^{1/p}}{r!\,n^{1/\gamma}} = \frac{\|\varphi\|_\gamma}{r!\,n^r}\,. \end{aligned}$$

□

Here are two simple examples:

1. (Rice [A-1969]). For $f(x) = x^\alpha$ on $[0,1]$, $\alpha > 0$, and $r = 1, 2, \dots$,

$$\sigma_{n,r}(f)_\infty \le C_{r,\alpha} n^{-r}, \qquad n = 1, 2, \dots . \tag{4.16}$$

This follows from Theorem 4.5 because $f^{(r)}$ is monotone and belongs to $L_{1/r}$.

2. For $f(x) = [1/\log(x/2)]^r$ on $[0,1]$, $r = 1, 2, \dots$,

$$\sigma_{n,r}(f)_\infty \le C_r n^{-r}, \qquad n = 1, 2, \dots . \tag{4.17}$$

Here we have $|f^{(r)}(x)| \le C|\log(x/2)|^{-r-1}x^{-r} =: \varphi(x)$ and $\varphi \in L_{1/r}$.

The first essential estimates of the approximation error $\sigma_{n,r}(f)_p$, better than those obtainable for fixed knot splines were found by Birman and Solomyak [1967] by a method now known as "adaptive approximation" which we briefly describe in the case $p = \infty$. It starts with the assumption that a desirable upper bound $\varepsilon > 0$ for the approximation error is given. If $f \in C(A)$, $A = [0,1]$, we divide A into two subintervals $[0, \frac12]$, $[\frac12, 1]$ of equal length. On each of them we find a polynomial of degree $\le r - 1$ of good approximation to f. If one of the errors will be $\le \varepsilon$, we are through with that subinterval. The remaining interval or intervals we again divide into parts of equal length, try to approximate f on each of them, and so on.

This will produce a certain number $N = N_\varepsilon(f)$ of disjoint (up to their endpoints) dyadic intervals of different lengths, covering [0,1]. On each of them, we will have a polynomial from $\mathcal{P}_{r-1}$, which approximates f with an error $\le \varepsilon$. Taken together, the polynomial pieces will generate a spline $S \in \Sigma_{rN,r}$ for which $\|f - S\|_\infty \le \varepsilon$.

Birman and Solomyak use this technique to prove error estimates for adaptive approximation which are only slightly worse than for free knot splines. For example, in Theorem 4.3, one replaces $\mathrm{Var}_A(f)$ by $\|f'\|_p$, $p > 1$. Because of its simplicity, adaptive approximation is very important in Numerical Analysis. It also extends readily to multivariate approximation.

§ 5. Smoothness in L_p for $0 < p < 1$

The improvement of free-knot spline as compared with fixed knot spline approximation manifests itself in a relaxation of the conditions on the approximated function f. If one wants to obtain, for the free knot error $\sigma_n(f)_p$ in $L_p(A)$, $A = [0,1]$, an estimate of the order (in some sense) $n^{-\alpha}$, $\alpha > 0$, one need only require, as we shall see in §8, smoothness properties of f in L_γ, $\gamma := (\alpha + \frac1p)^{-1}$ and not in L_p. The number γ (whose importance has been first recognized by Birman and Solomyak [1967]) is always less than p and often less than 1. This makes it necessary to study the properties of $\omega_r(f,t)_p$ and of Besov spaces $B^\alpha_q(L_p)$ for $p < 1$. While the results that we obtain are for the most part analogous to the case $p \ge 1$, the proofs are often more involved. This is at least in part due to the fact that there are no linear functionals on L_p, $0 < p < 1$. Therefore, results previously obtained by using linear operators such as the quasi-interpolants must now be proved in another way. A further minor difference is that $\|\cdot\|_p$ does not satisfy the triangle inequality. However, this can usually be circumvented by using one of the two inequalities, valid for $f, g \in L_p$, $0 < p \le \infty$:

$$\|f + g\|_p^\mu \le \|f\|_p^\mu + \|g\|_p^\mu, \tag{5.1}$$

where $\mu := \min(p, 1)$ and

$$\|f + g\|_p \le C(\|f\|_p + \|g\|_p), \tag{5.2}$$

with C depending only on p, and with $C = 1$ for $p \ge 1$.

For any real number t, let $T_t := T(t)$ be the translation operator $T(t)f := f(\cdot + t)$. If $A_t := [0, 1-t]$, $0 < t < 1$, we have $\|T(t)f - f\|_p(A_t) \to 0$, $t \to 0+$, for any $f \in L_p(A)$, $p > 0$. To see that this is true for $0 < p < 1$, we use the truncated functions $f_M(x) := \min(|f(x)|, M)\,\mathrm{sign}\, f(x)$, $M > 0$. They are bounded and satisfy $\|f - f_M\|_p \to 0$ as $M \to \infty$. Also, for $0 < p < 1$,

$$\|T(t)f_M - f_M\|_p(A_t) \le \|T(t)f_M - f_M\|_1(A_t) \to 0, \qquad t \to 0+ .$$

Since

$$\|T(t)f - f\|_p(A_t) \le C\{\|T(t)(f - f_M)\|_p(A_t) + \|T(t)f_M - f_M\|_p(A_t) + \|f_M - f\|_p(A_t)\},$$

the left side of this last inequality can be made arbitrarily small by first choosing M large and then $t > 0$ small. It follows that the modulus of smoothness $\omega_r(f,t)_p$, $t \ge 0$, defined by

$$\omega_r(f,t)_p := \sup_{0<h\le t} \left(\int_{A_{rh}} |\Delta_h^r(f,x)|^p\, dx \right)^{1/p} \tag{5.3}$$

is a continuous increasing function which vanishes at 0. In place of the subadditivity of ω_r, we have from (5.1)

$$\omega_r(f+g,t)_p^\mu \le \omega_r(f,t)_p^\mu + \omega_r(g,t)_p^\mu . \tag{5.4}$$

To derive further properties of ω_r, we shall need some identities for the translation and difference operators, $\Delta_h := \Delta(h)$. The latter are defined for each real h by $\Delta_h f(x) := f(x+h) - f(x)$. Then, $T_0 = I$, $\Delta_0 = 0$ and $\Delta_h = T_h - I$. If r numbers $h_1, \dots, h_r$ are given, then for each subset Λ of $\{1, \dots, r\}$ we write

$$\sigma_\Lambda := \sum_{j\in\Lambda} \frac{h_j}{j}, \qquad \tau_\Lambda := \sum_{j\in\Lambda} \frac{j-r}{j} h_j, \tag{5.5}$$

in particular $\sigma_\Lambda = \tau_\Lambda = 0$ if $\Lambda = \emptyset$. Then we have:

Lemma 5.1. *The following operator identities are valid:*

$$\Delta_{nh}^r = \Delta_h^r (I + T_h + \cdots + T_{(n-1)h})^r = \Delta_h^r \sum_{k=0}^{(n-1)r} a_k(n,r) T_{kh}, \tag{5.6}$$

where the coefficients a_k *satisfy* $0 \le a_k(n,r) \le n^{r-1}$, $k = 0, \dots, (n-1)r$, *and*

$$\prod_{j=1}^{r} \Delta(h_j) = \sum_{\Lambda} (-1)^{r-|\Lambda|} T(\tau_\Lambda) \Delta^r(\sigma_\Lambda) \tag{5.7}$$

with summation over all subsets Λ *of* $\{1, \dots, r\}$.

Proof. We have $\Delta_{h_1+h_2} = \Delta_{h_2} T_{h_1} + \Delta_{h_1}$. From this, we derive by induction on r, that $\Delta_{nh} = \Delta_h(I + T_h + \dots + T_{(n-1)h})$ and (5.6) follows. The inequalities for the a_k follow by induction.

To prove (5.7), we begin with the identity

$$\prod_{j=1}^{r} \Delta((j-k)h_j) = \prod_{j=1}^{r} [T((j-k)h_j) - I]$$
$$= \sum_{\Lambda} (-1)^{r-|\Lambda|} T\Big(\sum_{j\in\Lambda} j h_j\Big) T\Big(-\sum_{j\in\Lambda} k h_j\Big).$$

The left side is zero for $k = 1, \dots, r$. Hence, multiplying both sides by $(-1)^{r-k}\binom{r}{k}$ and summing over $k = 0, \dots, r$, we obtain

$$(-1)^r \prod_{j=1}^{r} \Delta(j h_j) = \sum_{\Lambda} (-1)^{r-|\Lambda|} T\Big(\sum_{j\in\Lambda} j h_j\Big) \Delta^r\Big(-\sum_{j\in\Lambda} h_j\Big).$$

Here we replace h_j by h_j/j, use the identity $\Delta^r(-\sigma_\Lambda) = (-1)^r T(-r\sigma_\Lambda)\Delta^r(\sigma_\Lambda)$ for each resulting term in the right sum and thereby arrive at (5.7). □

Useful inequalities follow from (5.6). Applying this identity to $f \in L_p$, we obtain with $A_s := [0, 1-s]$ and $0 < p \le 1$,

$$\|\Delta^r_{nh}(f)\|_p^p(A_{rnh}) \le \sum_{k=0}^{(n-1)r} a_k(n,r)^p \|\Delta^r_h(f)\|_p^p(A_{rh}) \le r n^{(r-1)p+1} \|\Delta^r_h(f)\|_p^p.$$

From this we derive

$$\omega_r(f, nt)_p \le r^{1/p} n^{r-1+\frac{1}{p}} \omega_r(f,t)_p, \qquad t > 0. \tag{5.8}$$

In particular, taking $n = [1/t] + 1$ in (5.8) we obtain for $f \in L_p$, $0 < p \le 1$, and some $C_r > 0$,

$$\omega_r(f,t)_p \ge C_r \omega_r(f,1)_p t^{r-1+\frac{1}{p}}, \qquad 0 < t \le 1. \tag{5.9}$$

This and the monotonicity of ω_r yield: *If* $\omega_r(f,t)_p = 0$ *for some particular* $t > 0$, *then* $\omega_r(f,t)_p = 0$ *identically on* $[0,1]$.

We next prove that if $\omega_r(f,t)_p = 0$ on $[0,1]$, then f coincides a.e. with some polynomial of degree $\le r-1$. We need the following lemma.

Lemma 5.2. *Let* $I = [a,b]$ *and* $0 < p \le 1$. *If* $f \in L_p(I)$, *then* (i) *there is a constant* α *such that*

$$\|f - \alpha\|_p(I) \le 2^{1/p} \omega(f, |I|, I)_p. \tag{5.10}$$

(ii) *Moreover,* f *is equal a.e. to a constant on* I *if and only if* $\Delta_s f(x) = 0$ *a.e. on* $I_s := [a, b-s]$ *for almost every* s, $0 < s \le b-a$.

Proof. We have

$$\int_a^b \int_a^b |f(y) - f(x)|^p \, dx dy = \int_a^b \int_x^b |f(y) - f(x)|^p \, dy dx + \int_a^b \int_a^x |f(y) - f(x)|^p \, dy dx. \tag{5.11}$$

Making the change of variables $y = x + s$ in the first and $y = x - s$ in the second integral and changing the orders of integration, we see that this is equal to

$$\int_0^{b-a} \int_a^{b-s} |f(x) - f(x+s)|^p \, dx ds + \int_0^{b-a} \int_{a+s}^{b} |f(x) - f(x-s)|^p \, dx ds = 2 \int_0^{b-a} \int_a^{b-s} |f(x+s) - f(x)|^p \, dx ds. \tag{5.12}$$

Comparing (5.11) and (5.12), and using the continuity s of the last interior integral we derive (ii). Also,

$$\frac{1}{b-a} \int_a^b \int_a^b |f(y) - f(x)|^p \, dx dy \le 2\omega(f, b-a, I)_p^p.$$

The left side is an average of a function of y over $[a,b]$. Consequently for some y and $\alpha := f(y)$,

$$\int_a^b |f(x) - \alpha|^p \, dx \le 2\omega(f, b-a, I)_p^p.$$ □

Theorem 5.3. *If for* $f \in L_p(A)$, $0 < p \le 1$, $r = 1, 2, \dots$, *one has* $\omega_r(f,t)_p = 0$ *for some* $0 < t \le 1$, *then* f *is equivalent (equal a.e.) on* A *to a polynomial of degree* $\le r-1$.

Proof. For continuous functions and $r = 1, 2, \dots$, this follows from Proposition 7.1 of Chapter 2. One sees easily that the assertion $\omega_r(f,t)_p = 0$ is identical with

$$\Delta^r_h f(x) = 0 \text{ for almost all pairs } x, h, \qquad 0 \le x < x + h \le 1. \tag{5.13}$$

(in the sense of two-dimensional measure). To complete the proof, we shall show by induction that (5.13) implies $f \in \mathcal{P}_r$. For $r = 1$, this follows from Lemma 5.2(ii).

We use formula (5.7) for $\Delta_s^{r-1}\Delta_t(f,x)$. With $h_1 = \cdots = h_{r-1} = s$ and $h_r = t$, it is applicable if all terms of the sum (5.7) are defined, that is, if all $x + \tau_\Lambda + r\sigma_\Lambda \le 1$. Given $\varepsilon > 0$, there is a constant $c = c(r,\varepsilon) < \varepsilon$, for which $\tau_\Lambda + r\sigma_\Lambda \le \varepsilon$ whenever $0 \le s,t \le c$. By the inductive assumption and (5.7), $\Delta_t(f,x)$ is a polynomial of degree $< r-1$:

$$f(x+t) - f(x) = \sum_{k=0}^{r-2} a_k(t)x^k \tag{5.14}$$

for almost all x, t with $x \in [0, 1-\varepsilon]$, $t \in [0,c]$.

We prove that the a_k are equivalent to continuous functions. Applying to (5.14) the operator Δ_h for the variable t we get, for all $0 \le h < \varepsilon$, and almost all $x \in [0, 1-2\varepsilon] =: J$, $0 \le t \le c$,

$$f(x+h+t) - f(x+t) = \sum_{k=0}^{r-2} \Delta_h(a_k,t)x^k\,.$$

By the equivalence of two (quasi-) norms on $\mathcal{P}_{r-2}$,

$$\begin{aligned}\max_{k=0,\dots,r-2} |\Delta_h(a_k,t)| &\le C\|f(\cdot+h+t) - f(\cdot+t)\|_p(J)\\ &\le C\|T_h f - f\|_p \to 0\,, \quad h \to 0\end{aligned}$$

for almost all t. Thus, $a_k(t)$ are uniformly continuous on a set of full measure of $[0,c]$ and admit a continuous extension.

Next we apply to (5.14) the operator Δ_h^r with respect to t. On the left we get $\Delta_h^r f(x+t) = T_t\Delta_h^r(f,x) = 0$, so that

$$\sum_{k=0}^{r-2} \Delta_h^r(a_k,t)x^k = 0 \text{ a.e. for } x \in J\,, \quad t \in [0,c]\,.$$

Hence $\Delta_h^r(a_k,t) = 0$, a.e., $0 < t \le c$. Since a_k is continuous, by Theorem 7.1 of Chapter 2, $a_k(t)$ is equivalent to a polynomial in $\mathcal{P}_{r-1}$. We fix an $x \in J$ for which (5.14) holds a.e. in $t \in [0,c]$. This shows that f is equal a.e. to a polynomial in $\mathcal{P}_{r-1}$ on $[x, x+c]$. Since the intervals $[x,x+c]$ overlap and cover J, the polynomials are all identical and $f = P$, a.e. on J, $P \in \mathcal{P}_{r-1}$. Since ε is arbitrary, the proof is complete. □

In this Chapter, we shall also occasionally use the averaged modulus of smoothness $w_r(f,t)_p$, $0 < p \le 1$. (For $1 \le p \le \infty$, a related averaged modulus has been given in (5.1), Chapter 6.) It is defined for $f \in L_p(I)$, $I = [a,b]$, $r = 1, 2, \dots$, $t > 0$ by

$$w_r(f,t)_p := w_r(f,t,I)_p := \left(\frac{1}{t}\int_0^t \|\Delta_s^r(f,\cdot)\|_p^p(I_{rs})\,ds\right)^{1/p}, \quad 0 < p \le 1\,, \tag{5.15}$$

where for any $\delta > 0$, $I_\delta := [a, b-\delta]$. We have a (restricted) subadditivity of w_r: If $I = \bigcup_{j=1}^n I_j$, where I_j are intervals with disjoint interiors, $j = 1, \dots, n$, then

$$\sum_{j=1}^n w_r(f,t,I_j)_p^p \le w_r(f,t,I)_p^p\,. \tag{5.16}$$

Indeed the left side is

$$\frac{1}{t}\int_0^t \sum_{j=1}^n \int_{(I_j)_{rh}} |\Delta_h^r(f,x)|^p\,dx\,dh \le \frac{1}{t}\int_0^t \int_J |\Delta_h^r(f,x)|^p\,dx\,dh$$

where $J := \cup(I_j)_{rh}$.

More generally, if the intervals I_j are not necessarily disjoint but a point $x \in I$ appears in at most C of the I_j, then (5.16) holds with the right side replaced by $Cw_r(f,t,I)_p^p$.

Another property of w_r is its equivalence to the ordinary modulus of smoothness ω_r. There are constants C_1, C_2 depending only on p and r such that for all $f \in L_p$, $0 < p < \infty$,

$$C_1 w_r(f,t)_p \le \omega_r(f,t)_p \le C_2 w_r(f,t)_p\,, \qquad 0 < t \le |I|\,. \tag{5.17}$$

The left inequality is obvious, and the right one is proved in an identical way to (5.6) of Chapter 6, except that for $0 < p < 1$ one uses inequality (5.1) in place of the triangle inequality. This establishes (5.17) for $0 \le t \le (4r)^{-1}|I|$. By using the inequalities $\omega_r(f,t)_p \le C\omega_r(f,(4r)^{-1}t)$ and $w_r(f,(4r)^{-1}t) \le (4r)^{rp}w(f,t)$ we obtain (5.17). As an application we prove the M. Riesz' theorem about precompact sets $\mathcal{F}$ in L_p, which is usually stated for $p \ge 1$.

Theorem 5.4. *Let $0 < p < \infty$, $r = 1, 2, \dots$, $M > 0$. Then a set $\mathcal{F}$ of functions $f \in L_p(A)$ is precompact in L_p if*

$$w_r(f,t)_p \to 0 \text{ for } t > 0 \text{ uniformly for } f \in \mathcal{F};\ \|f\|_p \le M\,. \tag{5.18}$$

Proof. We can assume that $A = [0,1]$. We can also assume that $r = 1$, for if $r \ge 2$, and for example $0 < p \le 1$, Marchaud's inequality (8.10) of Chapter 2 yields

$$\omega(f,t)_p^p \le Ct^p\left(\int_t^1 \frac{\omega_r(f,s)_p^p\,ds}{s^{p+1}} + M^p\right)$$

and this converges to zero for $t \to 0$, uniformly for $f \in \mathcal{F}$.

For $n = 1, 2, \dots$, we shall approximate $f \in \mathcal{F}$ by step functions $S_n \in S_1(T_n, A)$ with equally spaced knots $T_n := (t_j)_1^{n-1}$, $t_j := j/n$, $j = 0, \dots, n$. According to Lemma 5.2 and (5.17), for each of the intervals $I_j := (t_j, t_{j+1})$, there is a constant α_j satisfying

$$\|f - \alpha_j\|_p(I_j) \le C\omega_1\left(f,(4r)^{-1}|I_j|, I_j\right)_p \le C\omega_1\left(f, n^{-1}, I_j\right)_p\,. \tag{5.19}$$

Let $S(x) := \alpha_j$, for $x \in I_j$, $j = 1, \dots, n$. Raising each side of (5.19) to the power p and adding the resulting inequalities, we obtain from (5.16):

$$\|f - S\|_p(A) \le C w_1 \left(f, n^{-1}, A\right)_p \le C\omega \left(f, n^{-1}\right)_p . \tag{5.20}$$

For given $\varepsilon > 0$, it follows from (5.20) and (5.18) that we can choose n so that

$$\|f - S\|_p(A) < \varepsilon, \quad f \in \mathcal{F}.$$

From this and (5.18) it follows that the set $\mathcal{F}_1$ of all S that appear in (5.20) is bounded. It is precompact as a subset of a finite dimensional subspace $\mathcal{S}_1(T_n)$ of L_p. There is a finite ε-net for $\mathcal{F}_1$, hence a finite 3ε-net for $\mathcal{F}$, and therefore $\mathcal{F}$ is precompact in L_p. □

It is not difficult to prove that conditions (5.18) are also *necessary* for the precompactness of $\mathcal{F}$.

We next discuss approximation by algebraic polynomials in the metric of L_p, $0 < p < 1$. As usual, we define $E_n(f, A)_p := \min_{P \in \mathcal{P}_n} \|f - P\|_{L_p(A)}$. The existence of the minimum is standard, but there is no uniqueness here. For example if f takes the value -1 on $[-1, 0]$ and $+1$ on $[0, 1]$, then the two polynomials ± 1 are both best $L_p[-1, 1]$ approximations from $\mathcal{P}_0$ to f if $0 < p < 1$.

It is now simple to extend Theorem 4.2 of Chapter 6 of Whitney to the case $0 < p < 1$.

Theorem 5.5 (Storozhenko [1978]). *If $r = 1, 2, \dots$ and $0 < p < 1$, then for each $f \in L_p(A)$, $A = [0, 1]$,*

$$E_r(f, A)_p \le C\omega_r(f, |A|, A)_p \tag{5.21}$$

with the constant C depending only on r and p.

Proof. If (5.21) were not true, then for each $n = 1, 2, \dots$ there would exist an $f_n \in L_p$ with the properties $E_r(f_n, A)_p = \|f_n\|_p = 1$ and $\omega_r(f_n, |A|)_p \le 1/n$. Since $\omega_r(f_n, t)_p \to 0$ for each n, and moreover for all n and t, $\omega_r(f_n, t)_p \le 1/n$, this convergence is uniform in n. Thus, the set $\mathcal{F} := \{f_n, n = 1, 2, \dots\}$ satisfies (5.18) and hence is precompact. It contains a convergent subsequence $f_{n_k} \to f$ in L_p. Then $\omega_r(f, t)_p = 0$ for all $0 < t \le 1$ and by Theorem 5.3, $f = P$ a.e. for some $P \in \mathcal{P}_{r-1}$. But this is a contradiction in view of $E_r(f, A)_p = 1$. □

As an application of Whitney's theorem, we prove (Shalashova [1972]; see also Storozhenko [1978] and Shvedov [1979]) that the Jackson estimate for approximation by algebraic polynomials is valid in $L_p(A)$, $A = [0, 1]$, $0 < p < 1$.

Theorem 5.6. *If $r = 1, 2, \dots$ and $0 < p < 1$, then for each $f \in L_p(A)$, $A = [0, 1]$, and $n = r, r + 1, \dots,$*

$$E_n(f, A)_p \le C\omega_r(f, 1/n)_p \tag{5.22}$$

with the constant C depending only on r and p.

Proof. We fix r and $n \ge r$ and let $J_j := [\xi_{j-1}, \xi_j]$, $j = 1, \dots, n$ with $\xi_j := j/n$, $j \ge 0$. According to Whitney's theorem and (5.17), for each interval $I_j := [\xi_{j-1}, \xi_{j+2}] \cap A$, there is a polynomial $P_j \in \mathcal{P}_{r-1}$ which satisfies

$$\|f - P_j\|_p(I_j) \le C\,\omega_r(f, |I_j/4r|, I_j)_p \le C\, w_r(f, 1/n, I_j)_p . \tag{5.23}$$

Then using the subadditivity (5.16) of w_r, it follows that the piecewise polynomial

$$S(x) := P_j(x), \quad x \in J_j, \ j = 1, \dots, n, \tag{5.24}$$

satisfies

$$\|f - S\|_p(A) \le C_r w_r(f, 1/n, A)_p \le C_r \omega_r(f, 1/n)_p . \tag{5.25}$$

We next approximate S. We have with $R_j := P_j - P_{j-1}$, $j = 2, \dots, n$, $R_1 := P_1$,

$$S = R_1 + \sum_{j=2}^{n} R_j \varphi_j, \quad \varphi_j := \chi_{[\xi_j, 1]} . \tag{5.26}$$

We have shown in the derivation of (7.13) of Chapter 8 that there is a polynomial $Q_j \in \mathcal{P}_n$ such that

$$\begin{aligned} |\varphi_j(x) - Q_j(x)| &\le C \min(1, (n|x - \xi_j|)^{-4r+1}) \\ &\le C \min(1, (n|x - \xi_j|)^{-3r}) \end{aligned} \tag{5.27}$$

Therefore, the polynomial $P := R_1 + \sum_{j=2}^{n} R_j Q_j$ is in $\mathcal{P}_{n+r}$ and satisfies, for each $k = 1, \dots, n$,

$$\begin{aligned} \int_{J_k} |S(x) - P(x)|^p \, dx &\le C \sum_{j=2}^{n} \|R_j\|_\infty(J_k)^p \int_{J_k} \min(1, (n|x - \xi_j|)^{-3rp}) \, dx \\ &\le C \sum_{j=2}^{n} \|R_j\|_\infty(J_k)^p (1 + |j - k|)^{-3rp} \cdot 1/n . \end{aligned} \tag{5.28}$$

If $I_{j,k}$ is the smallest interval which contains I_j and I_k then from the polynomial inequalities (2.10) and (2.16) of Chapter 4, we have

$$\begin{aligned} \|R_j\|_\infty(I_{j,k}) &\le C\,(|j - k| + 1)^r \|R_j\|_\infty(J_j) \le C\, n^{1/p}(|j - k| + 1)^r \|R_j\|_p(J_j) \\ &\le C\, n^{1/p}(|j - k| + 1)^r \left[\|f - P_j\|_p(I_j) + \|f - P_{j-1}\|_p(I_{j-1})\right] \\ &\le C\, n^{1/p}(|j - k| + 1)^r \left[w_r(f, 1/n, I_j)_p + w_r(f, 1/n, I_{j-1})_p\right] . \end{aligned} \tag{5.29}$$

If we replace $\|R_j\|_\infty$ in (5.28) by the right side of (5.29), we obtain

$$\begin{aligned}\|S-P\|_p(A)^p &\le C\sum_{j=2,k=1}^{n,n}(|j-k|+1)^{-2rp}w_r(f,1/n,I_j)_p^p \\ &\le C\sum_{k=1}^{n}w_r(f,1/n,I_j)_p^p.\end{aligned} \tag{5.30}$$

By the remark following (5.16), the right side of (5.30) does not exceed $Cw_r(f,1/n,A)_p^p \le \omega_r(f,1/n)_p^p$. Therefore, (5.22) follows from (5.25) and (5.30). □

§ 6. Dyadic Splines in L_p, $0 < p < 1$

This section contains an extension of the theory of §3 to the spaces L_p with range $0 < p < 1$. New difficulties arise for $p < 1$; they will be overcome by means of the results of §5. In addition to their own interest, the dyadic splines will allow the characterization of the Besov spaces $B_q^\alpha(L_p)$, $0 < p < 1$, in §7. This will be the main tool in characterizing the free knot approximation spaces in §8.

Let $\Delta_n := (t_j)_1^{2^n-1}$ be the n-th dyadic knot sequence (3.1) in $A = [0,1]$; we let $t_j := j2^{-n}$, $j \le 0$, $j \ge 2^n$. The quasi-interpolants $Q_{n,r}$ of §4 of Chapter 5, used in §3 for $p \ge 1$, are not defined on L_p. To create a dyadic spline approximation to $f \in L_p$, $p < 1$, we shall first approximate f by a piecewise polynomial S and then take $Q_n(S)$ as the dyadic spline approximation to f. It will be important in the applications that follow to retain as much flexibility as possible in the choice of S. This will be accomplished by using the idea of "near best" approximation. If $\tau \ge 1$ and I is a subinterval of A, we say that $P \in \mathcal{P}_{r-1}$ is a *τ-near best approximation* to f in $L_p(I)$ if

$$\|f-P\|_p(I) \le \tau E_{r-1}(f,I)_p\,. \tag{6.1}$$

Now, we fix a value of $p_0 < 1$ and fix a value of τ. For each j, let $I_j := [t_{j-r+1}, t_{j+r}] \cap A$ and let $P_j \in \mathcal{P}_{r-1}$ be one of the polynomials which is a τ-near best L_{p_0} approximation to f on I_j. Then

$$S_n(f,x) := P_j(x)\,, \quad x \in (t_j, t_{j+1})\,, \quad j = 0, \dots 2^n - 1 \tag{6.2}$$

is a piecewise polynomial with breakpoints t_j, $j = 1, \dots, 2^n - 1$. For a fixed sequence of quasi-interpolants Q_n, $n = 0, 1, \dots, Q_{-1} = 0$ we define

$$\overline{Q}_n(f) := \overline{Q}_{n,r}(f) := Q_{n,r}(S_n)\,, \quad f \in L_{p_0}\,, \quad n = -1, 0, 1, \dots\,. \tag{6.3}$$

We thus obtain operators $\overline{Q}_{n,r}$ mapping L_{p_0} into $\mathcal{S}_r(\Delta_n, A)$, which will be substitutes for the $Q_{n,r}$.

Summing up, we have defined the operator $S_n(f)$ by (6.2); this S_n depends not only on n and r but also on τ, p_0 $(0 < p_0 \le \infty)$ and the particular choice

of near best approximant. We shall denote by $\mathcal{S}(n,r,\tau,p_0)$ the collection of all such S_n. The quasi-interpolant operators $Q_n := Q_{n,r}$ are from Chapter 5 and so $\overline{Q}_n$ depends on the same parameters as S_n and the choice of quasi-interpolant $Q_{n,r}$. Let $\mathcal{Q}(n,r,\tau,p_0)$ denote the collection of all such operators $\overline{Q}_n$. The error estimates we derive in this section for approximation apply to any $\overline{Q}_n \in \mathcal{Q}(n,r,\tau,p_0)$ with constants that depend only on p_0, r, and τ.

The following properties of near best polynomials while interesting in their own right will also be used to establish the approximation properties of the $\overline{Q}_n$ (Theorem 6.3) and lead to interpolation theorems for Besov spaces in the next section.

Lemma 6.1. *There is a constant $C = C(p_0, r)$ with the following properties. Let $f = g_1 + g_2$, $g_i \in L_p(I)$, $i = 1, 2$, $p \ge p_0$. Let $P \in \mathcal{P}_{r-1}$ be a polynomial of τ-near best approximation to $f \in L_p$, then there exist polynomials $P_i \in \mathcal{P}_{r-1}$, which are $C\tau$-near best for g_i, $i = 1, 2$, and satisfy $P = P_1 + P_2$.*

Proof. We write $E(h) := E_{r-1}(h, I)_p$ and omit the subscript p. We can assume that $E(g_1) \le E(g_2)$. Let P_1 be the polynomial of best approximation to g_1 and let $P_2 := P - P_1$. Then with the constant C of the quasi-normed space L_{p_0},

$$\begin{aligned}\|g_2 - P_2\| = \|(f-P) - (g_1 - P_1)\| &\le C\big(\tau E(f) + E(g_1)\big) \\ &\le C^2\tau\big(E(g_1) + E(g_2)\big) + CE(g_1) \le C_1\tau E(g_2)\,,\end{aligned}$$

where $C_1 = 2C^2 + C$. □

For the next lemma, we need the inequality (2.16) of Chapter 4, which asserts that for $0 < p_0 < p_1 \le \infty$, $\theta = \frac{1}{p_1} - \frac{1}{p_0}$, $P \in \mathcal{P}_{r-1}$, we have $\|P\|_{p_1}(I) \le C|I|^\theta \|P\|_{p_0}(I)$, with C depending on r and p_0.

Lemma 6.2. *There is a constant $C := C(r, p_0)$ with the following property. If $f \in L_{p_1}(I)$, $0 < p_0 < p_1 \le \infty$, and if $P \in \mathcal{P}_{r-1}$ is a near τ-best L_{p_0} approximation to f on I, then it is also a $C\tau$-near best L_{p_1} approximation to f on I.*

Proof. Let P^* be a polynomial of best $L_{p_1}(I)$ approximation to f. Then on I,

$$\begin{aligned}\|f-P\|_{p_1} &\le C\big(E_{r-1}(f)_{p_1} + \|P^*-P\|_{p_1}\big) \\ &\le CE_{r-1}(f)_{p_1} + C|I|^\theta\|P^*-P\|_{p_0} \\ &\le CE_{r-1}(f)_{p_1} + C|I|^\theta\big(\|f-P^*\|_{p_0} + \|f-P\|_{p_0}\big) \\ &\le CE_{r-1}(f)_{p_1} + (\tau+1)C|I|^\theta\|f-P^*\|_{p_0} \\ &\le CE_{r-1}(f)_{p_1} + C(\tau+1)\|f-P^*\|_{p_1} \le C\,\tau E_{r-1}(f)_{p_1}\,.\end{aligned}$$

In this chain of inequalities, the second depends on Nikolskii's relation, the fifth is obtained by means of Hölder's inequality for $f - P^*$ on I. □

Let $\overline{Q}_n$ denote any of the operators in the class $\mathcal{Q}(n,r,\tau,p_0)$. Although the operators $\overline{Q}_n$ are not linear, they possess many of the approximation properties of the Q_n. From Whitney's theorem (5.21) and (5.17), it follows that for $p_0 \le p \le \infty$,

$$\|f - P_j\|_p[t_j, t_{j+1}] \le C\|f - P_j\|_p(I_j) \le Cw_r(f, 2^{-n}, I_j)_p\,, \quad f \in L_p$$

where w_r is the averaged modulus of smoothness (5.15). Since a point $x \in A$ is in at most $2r$ of the intervals I_j, we can add these estimates (see the remarks following (5.16)) and obtain for any of the piecewise polynomials $S_n(f)$ of (6.2),

$$\|f - S_n(f)\|_p(A) \le Cw_r(f, 2^{-n}, A)_p\,. \tag{6.4}$$

In order to replace here S_n by $\overline{Q}_n$, we first prove that $\overline{Q}_n$ is bounded on $L_p(A)$:

$$\|\overline{Q}_{n,r}(f)\|_p \le C\|f\|_p\,, \quad p \ge p_0\,, \quad C = C(r, p_0)\,. \tag{6.5}$$

Indeed, for each $f \in L_p$, by Lemma 6.2

$$\begin{aligned}\|S_n(f) - f\|_p^p &\le \sum_{j=0}^{2^n-1} \int_{t_j}^{t_{j+1}} |S_n(f,x) - f(x)|^p\,dx \le \sum_{j=0}^{2^n-1} \int_{I_j} |P_j - f|^p\,dx \\ &\le C \sum_{j=0}^{2^n-1} \int_{I_j} |f|^p\,dx \le C\|f\|_p^p\,,\end{aligned}$$

and it follows that $\|S_n(f)\|_p \le C\|f\|_p$.

We write $\overline{Q}_n(f)$ in its B-spline series: $\overline{Q}_n(f) = Q_n(S_n) = \sum_{k \in \Lambda_n} \gamma_k(S_n) N_k$ with $\Lambda_n := \{j : -r < j < 2^n\}$. Here γ_k is an extension of the coefficient functional c_k (4.1), Chapter 5. We have $c_k(S_n) = \gamma_k(S_n)$ because S_n is a polynomial on $[t_k, t_{k+1}]$. Therefore, from (4.6) of Chapter 5, we have

$$|\gamma_k(S_n)| \le C2^{-n/p}\|S_n\|_p([t_k, t_{k+r}] \cap A)\,, \qquad k \in \Lambda_n\,. \tag{6.6}$$

Moreover, (4.9) of Chapter 5 and (6.6) give

$$\begin{aligned}\|Q_n(S_n)\|_p &\le C\left(\sum_{k \in \Lambda_n} |\gamma_k(S_n)|^p |t_{k+r} - t_k|\right)^{1/p} \\ &\le C\|S_n\|_p\,, \quad p_0 < p \le \infty\,,\end{aligned} \tag{6.7}$$

because a point $x \in A$ appears in at most r of the intervals $[t_k, t_{k+r}] \cap A$. This establishes (6.5).

We can now apply Theorem 4.5 of Chapter 5 to obtain for $p \ge p_0$,

$$\|S_n - Q_n(S_n)\|_p \le C \begin{cases} \left(\displaystyle\sum_{0 \le j < 2^n} E_r(S_n, I_j)_p^p\right)^{1/p}, & p_0 \le p < \infty\,, \\ \displaystyle\max_{0 \le j < 2^n} E_r(S_n, I_j)_p\,, & p = \infty\,, \end{cases} \tag{6.8}$$

with C depending only on r and p_0.

With these remarks, we can now prove

Theorem 6.3. *Let $\tau \ge 1$ and $p_0 \le p \le \infty$. For each $f \in L_p$, and each operator $\overline{Q}_n$ in the class $\mathcal{Q}(n,r,\tau,p_0)$, we have*

$$d_n(f)_p \le \|f - \overline{Q}_n(f)\|_p \le C\omega_r(f, 2^{-n})_p \tag{6.9}$$

with C depending only on τ, r and p_0.

Proof. We consider one of the intervals I_j, $0 \le j < 2^n$, and suppose that $J := [t_i, t_{i+1}] \subset I_j$. Then

$$\begin{aligned}\|S_n - P_j\|_p(J) = \|P_i - P_j\|_p(J) &\le C\big(\|f - P_i\|_p(J) + \|f - P_j\|_p(J)\big) \\ &\le C\big(E_{r-1}(f, I_i)_p + E_{r-1}(f, I_j)_p\big) \le CE_{r-1}(f, I_j^*)_p\end{aligned} \tag{6.10}$$

where $I_j^* := \cup\{I_i : I_i \cap I_j \ne \emptyset\}$. If we raise both sides of (6.10) to the power p and sum over all $J \subset I_j$, we find for each j

$$E_{r-1}(S_n, I_j)_p \le CE_{r-1}(f, I_j^*)_p\,.$$

Now $|I_j^*| \le 6r2^{-n}$ and so Whitney's theorem yields

$$E_{r-1}(S_n, I_j)_p \le Cw_r(f, 2^{-n}, I_j^*)_p\,. \tag{6.11}$$

Let $p < \infty$. Since a point $x \in A$ appears in at most $6r$ of the intervals I_j^*, we can raise both sides of (6.11) to the power p, add the resulting terms, and use the remark following (5.16) to obtain

$$\sum_{0 \le j < 2^n} E_{r-1}(S_n, I_j)_p^p \le Cw_r(f, 2^{-n}, A)_p^p\,. \tag{6.12}$$

Substituting this into (6.8), we obtain $\|S_n - Q_n(S_n)\|_p \le Cw_r(f, 2^{-n}, A)_p$. This, (6.4) and the relation $w_r(f, 2^{-n}, A)_p \le C\omega_r(f, 2^{-n})_p$ establish (6.9). When $p = \infty$, a similar argument applies. □

We can also prove inverse estimates.

Theorem 6.4. *Let $0 < p \le \infty$, $r = 1, 2, \ldots$ and $\lambda := r - 1 + 1/p$. If $f \in L_p$ and $n = 1, 2, \ldots$, then*

$$\omega_r(f, 2^{-n})_p \le C2^{-n\lambda}\left(\sum_{k=0}^{n} 2^{k\lambda\mu} d_k(f)_p^\mu\right)^{1/\mu}, \qquad \mu := \min(1, p)\,. \tag{6.13}$$

Proof. When $p \ge 1$, this is Theorem 3.2; when $p < 1$, the same proof applies with (5.1) used in place of the triangle inequality. □

As in §3, we can use the companion inequalities (6.9) and (6.13) to characterize membership in the Besov spaces.

Theorem 6.5. *Let $r = 1, 2, \dots$, $0 < p \le \infty$, and $\lambda := r - 1 + 1/p$. If $0 < q \le \infty$ and $0 < \alpha < \min\{\lambda, r\}$, then $|f|_{B_q^\alpha(L_p)}$ is equivalent to*

$$(6.14)\qquad \begin{cases} \left(\sum_{n=0}^{\infty} [2^{n\alpha}\varepsilon_n(f)_p]^q\right)^{1/q}, & 0 < q < \infty \\ \sup_{n\ge 0} 2^{n\alpha}\varepsilon_n(f)_p, & q = \infty \end{cases}$$

where $\varepsilon_n(f)_p$ can be any of the quantities: $d_n(f)_p$, $\|f - \overline{Q}_n(f)\|_p$, $\|\overline{Q}_{n+1}(f) - \overline{Q}_n(f)\|_p$, for any of the operators $\overline{Q}_n$. Here, the constants of equivalency depend only on τ, r, p_0, q and α.

Proof. From (6.8), we have $\varepsilon_n(f)_p \le C\omega_r(f, 2^{-n})_p$ and therefore (6.14) is by (10.5) of Chapter 2 majorized by $|f|_{B_q^\alpha(L_p)}$. The required inequality in the opposite direction for the first and second choices of $\varepsilon_n(f)_p$ follows from (6.13) and the discrete Hardy inequality (1.7). For the third choice, we apply in addition the inequality

$$d_n(f)_p \le C\left(\sum_{k=n}^{\infty} \varepsilon_k(f)_p^\mu\right)^{1/\mu}, \qquad \mu := \min(1, p)$$

as in the derivation of (3.14) from (3.13). □

§ 7. Comparison of the Spaces $l_q^\alpha(L_p)$ and $B_q^\alpha(L_p)$

The spaces $l_q^\alpha(L_p)$ for $0 < p, q \le \infty$, $\alpha > 0$ have been defined in §7 of Chapter 6. They consist of all sequences $\boldsymbol{a} = (a_n)_0^\infty$ of functions $a_n \in L_p(A)$ for which the following (quasi-) norm is finite

$$(7.1)\qquad \|\boldsymbol{a}\|_{l_q^\alpha(L_p)} := \left(\sum_{n=0}^{\infty} [2^{n\alpha}\|a_n\|_p]^q\right)^{1/q}, \qquad 0 < q < \infty,$$

(with a proper supremum in place of the sum for $q = \infty$).

With the help of the spaces $l_q^\alpha(L_p)$ we can reformulate Theorem 6.5. Let $\tau \ge 1$, $p_0 > 0$, $r = 1, 2 \dots$. For each $n = 0, 1, \dots$, we let $\overline{Q}_n$ be one of the operators from the class $\mathcal{Q}(n, r, \tau, p_0)$ of the previous section: $\overline{Q}_n(f) = Q_n(S_n)$. We then define $\mathbf{R}f := (R_n(f))_0^\infty$ with $R_n(f) := \overline{Q}_n(f) - \overline{Q}_{n-1}(f)$, $n = 0, 1, \dots$, $\overline{Q}_{-1}(f) := 0$. If $0 < q \le \infty$, $p_0 \le p \le \infty$, and $\alpha < r - 1 + 1/p$, then the following (quasi)norms are equivalent:

$$(7.2)\qquad \|f\|_{B_q^\alpha(L_p)} \sim \|\mathbf{R}f\|_{l_q^\alpha(L_p)} \sim \|((d_n f)_p)_{-1}^\infty\|_{\alpha,q}$$

where $d_{-1}(f)_p := \|f\|_p$ and the constants of equivalency depend only on r, τ, p_0, and α. Here, we use the definition (3.10) of Chapter 2 for the α, q norm of a sequence.

The proof of (7.2) is quite simple. Theorem 6.5 gives that (7.2) holds with norms replaced by semi-norm (the terms R_0 and d_{-1} are deleted in the semi-norms) and therefore the first and last norms of (7.2) are equivalent. Also, since $\|R_0(f)\|_p \le C\|f\|_p$, the middle norm of (7.2) is less than a multiple of the first. On the other hand, $f = \sum_{n=0}^\infty R_n(f)$ with convergence in $\|\cdot\|_p$. Therefore, if $\mu < 1, p, q$, then Hölder's inequality gives

$$\|f\|_p \le \left(\sum_{n=0}^{\infty} \|R_n(f)\|_p^\mu\right)^{1/\mu} \le C\left(\sum_{n=0}^{\infty} [2^{n\alpha}\|R_n(f)\|_p]^q\right)^{1/q}.$$

Therefore the midddle norm in (7.2) is equivalent to the first.

In what follows the $\overline{Q}_n$ are arbitrary but fixed operators from $\mathcal{Q}(n, r, \tau, p_0)$ and $\mathbf{R}$ is obtained from the $\overline{Q}_n$ as above. The use of the spaces $l_q^\alpha(L_p)$ in their relation to $B_q^\alpha(L_p)$ given in the next theorem is similar to the *method of retracts*, discussed extensively in the book of Bergh and Löfström [B-1976]; however, the nonlinearity of the operators $\overline{Q}_n$ requires more care. We consider two pairs of spaces

$$(7.3)\qquad B_i := B_{q_i}^{\alpha_i}(L_{p_i}), \quad \tilde{l}_i := l_{q_i}^{\alpha_i}(L_{p_i}), \qquad i = 0, 1,$$

(where we assume that $p_0 \le p_1$), and have the result of DeVore and Popov [1988_2]:

Theorem 7.1. *Let $\alpha_i > 0$, $0 < q_i \le \infty$, $0 < p_0 \le p_1 \le \infty$, $r - 1 > \alpha_0, \alpha_1$ and $\tau \ge 1$. If $\overline{Q}_n$ and $\mathbf{R}$ are defined as above, then with constants $C_1, C_2 > 0$ that depend only on τ, p_0, α_0, α_1, q_0, q_1, r, one has for all $f \in B_0 + B_1$,*

$$(7.4)\qquad C_1 K(f, t; B_0, B_1) \le K(\mathbf{R}f, t; \tilde{l}_0, \tilde{l}_1) \le C_2 K(f, t; B_0, B_1).$$

Proof. (i) We prove the left inequality of (7.4). For a function $f \in L_{p_0}$, we assume that the sequence $\mathbf{R}f$ has a representation $\mathbf{R}f = \boldsymbol{a}^0 + \boldsymbol{a}^1$, $\boldsymbol{a}^i \in \tilde{l}_i$, $i = 0, 1$. Let $g_n := \overline{Q}_n(a_n^1)$ and $g := \sum_{j=0}^\infty g_j$. We first verify that $g \in L_{p_1}$. From (6.5), we have

$$\|g_n\|_{p_1} \le C\|a_n^1\|_{p_1}, \quad n = 0, 1, \dots .$$

Hence, for each $0 \le n < m < \infty$, and $\mu := \min(1, p_1)$, we have

$$(7.5)\qquad \left\|\sum_{j=n}^{m} g_j\right\|_{p_1} \le C\left(\sum_{j=n}^{m} \|a_j^1\|_{p_1}^\mu\right)^{1/\mu}.$$

Since $\boldsymbol{a}^1$ is in $\tilde{l}_1$, the right side of (7.5) tends to 0 as $n, m \to \infty$. This shows that the series defining g converges in the L_{p_1} norm and $g \in L_{p_1}$.

We shall estimate the K-functional on the left side of (7.4) by using the decomposition $f = f - g + g$. Since $\sum_{j=0}^{n} g_j$ is in $\mathcal{S}_r(\Delta_n)$, we have from (7.5), for $\mu := \min(1, p_1)$,

$$(7.6)\qquad d_n(g)_{p_1} \le \Big\| \sum_{j=n+1}^{\infty} g_j \Big\|_{p_1} \le C \Big(\sum_{j=n+1}^{\infty} \|a_j^1\|_{p_1}^{\mu} \Big)^{1/\mu}, \qquad n = -1, 0, \dots .$$

To (7.6), we can apply the Hardy inequality for the case (1.6)(i) and use the equivalence of Besov space norms (7.2) to obtain

$$(7.7)\qquad \|g\|_{B_1} \le C \|(d_n(g))\|_{\alpha_1, q_1} \le C \|(\|a_j^1\|_{p_1})\|_{\alpha_1, q_1} \le C \|\boldsymbol{a}^1\|_{l_1} .$$

We can prove a similar estimate for $f - g = \sum_{j=0}^{\infty} (R_j - g_j)$. Since $\sum_{j=0}^{n} (R_j - g_j) \in \mathcal{S}_r(\Delta_n)$, for $\mu := \min(1, p_0)$, we have

$$(7.8)\qquad d_n(f-g)_{p_0} \le \Big\| \sum_{j=n+1}^{\infty} (R_j - g_j) \Big\|_{p_0} \le \Big(\sum_{j=n+1}^{\infty} \|R_j - g_j\|_{p_0}^{\mu} \Big)^{1/\mu} .$$

Since $R_j := R_j(f) \in \mathcal{S}_r(\Delta_j)$, it is a piecewise polynomial of degree $< r$ with breakpoints Δ_j. It follows that $S_j(a_j^1) - R_j$ is a τ-near best L_{p_0} approximation to $a_j^1 - R_j$ among all such piecewise polynomials and therefore $\|S_j(a_j^1) - R_j\|_{p_0} \le C \|a_j^1 - R_j\|_{p_0}$. Using (6.7), and that Q_j is a projection onto $\mathcal{S}_r(\Delta_j)$, we obtain

$$\begin{aligned} \|R_j - g_j\|_{p_0} &= \|R_j - Q_j(S_j(a_j^1))\|_{p_0} = \|Q_j(R_j - S_j(a_j^1))\|_{p_0} \\ &\le C \|R_j - S_j(a_j^1)\|_{p_0} \le C \|R_j - a_j^1\|_{p_0} = C \|a_j^0\|_{p_0} . \end{aligned}$$

If we use this in (7.8) and argue as in the derivation of (7.7), we obtain

$$\|f - g\|_{B_0} \le C \|\boldsymbol{a}^0\|_{\tilde{l}_0} .$$

This inequality together with (7.7) gives

$$\|f - g\|_{B_0} + t\|g\|_{B_1} \le C(\|\boldsymbol{a}^0\|_{\tilde{l}_0} + t\|\boldsymbol{a}^1\|_{\tilde{l}_1}) .$$

Since $\mathbf{R}f = \boldsymbol{a}^0 + \boldsymbol{a}^1$ was an arbitrary decomposition, we derive the left inequality of (7.4).

(ii) To prove the right inequality of (7.4), we assume that a function $f \in B_0 + B_1 \subset L_{p_0}$ has the respresentation $f = g_0 + g_1$, $g_0 \in B_0$, $g_1 \in B_1$. Then $S_n := S_n(f)$ is a piecewise polynomial of order r with breakpoints Δ_n. By the definition of S_n, each polynomial piece P_j of S_n is a τ-near best approximant to f (on the interval I_j of §6). Applying Lemma 6.1 to each P_j, we find $P_j = P_j^0 + P_j^1$ with P_j^i $C\tau$-near best approximations in L_{p_0} to g_i on I_j, $i = 0, 1$. Let $S_n^i := P_j^i$ on (t_j, t_{j+1}). We define now a new operator $\tilde{S}_n$ which is in the class $\mathcal{S}(n, r, C\tau, p_0)$. It has the value $\tilde{S}_n(g_i) := S_{n_j}^i$, $i = 0, 1$. For other functions g it takes values in an arbitrary way as long as $\tilde{S}_n \in \mathcal{S}(n, r, C\tau, p_0)$. Then $\tilde{Q}_n := Q_n(\tilde{S}_n(\cdot)) \in \mathcal{Q}(n, r, C\tau, p_0)$. We define $\tilde{\mathbf{R}}$ from the $\tilde{Q}_n$ in the usual way. We then have $\mathbf{R}f = \tilde{\mathbf{R}}f = \tilde{\mathbf{R}}g_0 + \tilde{\mathbf{R}}g_1$. Now, (7.2) applies to $\tilde{\mathbf{R}}$ as well. Therefore, $\|\tilde{\mathbf{R}}g_i\|_{\tilde{l}_i} \le C\|g_i\|_{B_i}$, $i = 0, 1$, and

$$\|\tilde{\mathbf{R}}g_0\|_{\tilde{l}_0} + t\|\tilde{\mathbf{R}}g_1\|_{\tilde{l}_1} \le C(\|g_0\|_{B_0} + t\|g_1\|_{B_1}) .$$

The decomposition $f = g_0 + g_1$ was arbitrary, and from the definition of the K-functionals we derive the right inequality in (7.4). □

Using the definition of the θ, q-interpolation spaces in §7 of Chapter 6 we obtain from (7.4) for $\theta > 0$, $0 < q \le \infty$,

$$(7.9)\qquad \begin{cases} f \in (B_0, B_1)_{\theta,q} \ \text{ is equivalent to } \ \mathbf{R}f \in (l_0, l_1)_{\theta,q} , \\ \|f\|_{(B_0,B_1)_{\theta,q}} \sim \|\mathbf{R}f\|_{(l_0,l_1)_{\theta,q}} . \end{cases}$$

We now apply Theorem 7.5 of Chapter 6, which for $X = L_p$ asserts that

$$(7.10)\qquad (l_{q_0}^{\alpha_0}(L_p), l_{q_1}^{\alpha_1}(L_p))_{\theta,q} = l_q^\alpha(L_p)$$

whenever $0 < q \le \infty$ and $\alpha = (1-\theta)\alpha_0 + \theta\alpha_1$. Therefore, (7.2), (7.9), and (7.10) give that $\|f\|_{(B_0,B_1)_{\theta,q}}$ is equivalent to $\|f\|_{B_q^\alpha(L_p)}$. We have proved:

Theorem 7.2. *If $0 < \alpha_0, \alpha_1$, if $0 < \theta < 1$, if $0 < p, q_0, q_1 \le \infty$, if $0 < q \le \infty$, and if $\alpha = (1-\theta)\alpha_0 + \theta\alpha_1$, then*

$$(7.11)\qquad (B_{q_0}^{\alpha_0}(L_p), B_{q_1}^{\alpha_1}(L_p))_{\theta,q} = B_q^\alpha(L_p) .$$

If we take $X_0 = L_{p_0}$, $X_1 = L_{p_1}$ with $p_0 < p_1$ in Theorem 7.5 of Chapter 6, we obtain a formula similar to (7.10) with the restriction

$$(7.12)\qquad \frac{1}{q} = (1-\theta)\frac{1}{q_0} + \theta\frac{1}{q_1}, \qquad \alpha = (1-\theta)\alpha_0 + \theta\alpha_1 .$$

and with $l_q^\alpha((L_{p_0}, L_{p_1})_{\theta,q})$ on the right. This intermediate space is a Lorentz space $L_{p,q}$, namely (see **2** of §7, Chapter 2),

$$(L_{p_0}, L_{p_1})_{\theta,q} = L_{p,q} , \qquad \frac{1}{p} = (1-\theta)\frac{1}{p_0} + \theta\frac{1}{p_1} .$$

In the special case $q = p$ we have $L_{p,q} = L_p$ so that, with the same proof as before, we have

Theorem 7.3. *Let $\alpha_0, \alpha_1 > 0$, let $0 < \theta < 1$, let $0 < p_0, p_1, q_0, q_1 \le +\infty$, let α, q be defined by (7.12) and let $\frac{1}{p} = (1-\theta)\frac{1}{p_0} + \theta\frac{1}{p_1}$. If $p = q$, then*

$$(7.13)\qquad (B_{q_0}^{\alpha_0}(L_{p_0}), B_{q_1}^{\alpha_1}(L_{p_1}))_{\theta,p} = B_p^\alpha(L_p) .$$

§ 8. Free Knot Spline Approximation in L_p, $0 < p < \infty$

This section opens with the Jackson and Bernstein theorems (in the terminology of §5, Chapter 7) for this approximation by Petrushev. It is not very difficult to obtain from this a representation of the approximation spaces $A_q^\alpha(L_p)$ (defined in (1.8)) as a θ, q-intermediate space for L_p and some Besov space (see (8.16)). The main problem is then to identify these interpolation spaces. This has been achieved by DeVore and Popov for some values of the parameters, by means of §7. The approximation spaces turn out again to be some Besov spaces.

First results for free knot spline approximation have been by Peetre and Sparr [1972], by Burchard [1977]; later Brudnyi, in the related field of rational approximation, announced results of the type given in this section (but his proofs were never published). The older theorems for approximation by free knot splines identify approximation spaces $A_q^\alpha(L_p)$ as interpolation spaces of the form $(L_p, Y)_{\theta,q}$, often with a complicated space Y, different from one given here.

To characterize the approximation spaces $A_q^\alpha(L_p)$, the main problem is to find largest possible spaces for which some given estimate, such as $\sigma_n(f) \le Cn^{-\alpha}$, $n = 1, 2, \dots$ is valid. We shall see that the difference between free and fixed knot spline approximation is very large in this respect.

The theorems that we prove do not extend to all values of the parameters. Most notably, the case of $p = \infty$, which corresponds to uniform approximation, is missing. While the approximation spaces in this case can still be characterized as interpolation spaces, there is in general no explicit characterization of these interpolation spaces.

We begin with the embedding Theorem 8.1 of Besov spaces into L_p. Embedding theorems of this type are analogues of the well known Sobolev embedding theorem. The following elementary approach of Oswald [1980] to these embedding theorems is based on dyadic splines. We first note that for dyadic splines $S \in \mathcal{S}_r(\Delta_n, A)$, $A := [0, 1]$, we have, for $\alpha > 0$,

$$\|S\|_p \le C2^{n\alpha}\|S\|_\gamma, \tag{8.1}$$

where γ is the Sobolev embedding number

$$\gamma := \gamma(\alpha) := (\alpha + 1/p)^{-1}. \tag{8.2}$$

Indeed, on each dyadic interval $J_j = [t_j, t_{j+1}]$ with knots from Δ_n, S is a polynomial of degree $< r$. Hence, from (2.16) of Chapter 4,

$$\|S\|_p(J_j) \le C|J_j|^{1/p-1/\gamma}\|S\|_\gamma(J_j) = C2^{n\alpha}\|S\|_\gamma(J_j).$$

Therefore,

$$\begin{aligned}\|S\|_p^p &\le C2^{n\alpha p} \sum_{0\le j<2^n} \|S\|_\gamma^p(J_j) \\ &\le C2^{n\alpha p}\Big(\sum_{0\le j\le 2^n} \|S\|_\gamma^\gamma(J_j)\Big)^{p/\gamma} = C2^{n\alpha p}\|S\|_\gamma^p\end{aligned}$$

because $\gamma < p$, and the l_p norm does not exceed the l_γ norm.

Theorem 8.1. *Let $\alpha > 0$, $0 < p < \infty$, let $I = [a, b]$ and let γ be given by (8.2). Then $B_p^\alpha(L_\gamma(I))$ is continuously embedded in $L_p(I)$: the inequality*

$$\|f\|_p(I) \le C\|f\|_{B_p^\alpha(L_\gamma(I))} \le C\|f\|_{B_\gamma^\alpha(L_\gamma(I))} \tag{8.3}$$

holds for all $f \in B_p^\alpha(L_\gamma(I))$. In addition, if $r > \alpha$,

$$E_{r-1}(f, I)_p \le C|f|_{B_p^\alpha(L_\gamma(I))} \le C|f|_{B_\gamma^\alpha(L_\gamma(I))}. \tag{8.4}$$

Proof. The case of an arbitrary interval I reduces to the case $I = A = [0, 1]$ by linear substitution. Since $\gamma < p$, from (10.6) of Chapter 2, we have $|f|_{B_p^\alpha(L_\gamma)} \le C|f|_{B_\gamma^\alpha(L_\gamma)}$ and therefore it is sufficient to prove the first inequalities in (8.3) and (8.4). We take $r := [\alpha] + 1$ and for an $f \in L_p(A)$, the $R_n(f)$ of (7.2) corresponding to this r. Then for $\mu = \min(1, p)$, using (7.2) and (8.2),

$$\begin{aligned}\|f\|_p \le \Big(\sum_{n=0}^\infty \|R_n\|_p^\mu\Big)^{1/\mu} &\le C\Big(\sum_{n=0}^\infty (2^{n\alpha}\|R_n\|_\gamma)^\mu\Big)^{1/\mu} \\ &\le C\|f\|_{B_\mu^\alpha(L_\gamma)}.\end{aligned} \tag{8.5}$$

This is (8.3) if $p \le 1$. For $p > 1$, (8.5) yields $\|f\|_p \le C\|f\|_{B_1^\alpha(L_\gamma)}$. To improve this, we choose $1 < p_0 < p < p_1 < \infty$, with p_0, p_1 sufficiently close to p so that $\alpha_i := 1/\gamma - 1/p_i$, $i = 0, 1$ are positive. Then (8.5) gives

$$\|f\|_{p_i} \le C\|f\|_{B_1^{\alpha_i}(L_{p_i})}, \quad i = 0, 1. \tag{8.6}$$

We now interpolate between these two inequalities as follows. Let $0 < \theta < 1$ be chosen so that $\alpha = (1-\theta)\alpha_0 + \theta\alpha_1$. Then, it follows from the definition of the α_i, $i = 0, 1$ that $1/p = (1-\theta)/p_0 + \theta/p_1$ and $(L_{p_0}, L_{p_1})_{\theta,p} = L_p$. Hence, by (7.11), $(B_1^{\alpha_0}(L_\gamma), B_1^{\alpha_1}(L_\gamma))_{\theta,p} = B_p^\alpha(L_\gamma)$. Now, the inequality (8.6) says that the identity operator is a bounded mapping from $B_1^{\alpha_i}(L_\gamma)$ into L_{p_i}, $i = 0, 1$. Hence, by interpolation, the identity operator is bounded as a mapping from $B_p^\alpha(L_\gamma)$ into L_p which is (8.3).

To prove (8.4), we let $r > \alpha$ be arbitrary and apply (8.3) to $f - P$, where P is one of the best $L_\gamma(A)$ approximation polynomials of f of degree $< r$. Then we get

$$E_{r-1}(f, I)_p \le C\|f - P\|_{B_p^\alpha(L_\gamma)} = C\Big(\|f - P\|_\gamma + |f - P|_{B_p^\alpha(L_\gamma)}\Big).$$

Here $f - P$ and f have the same semi-norm; by Whitney's Theorem 5.5, $\|f - P\|_\gamma \le C\omega_r(f, 1)_\gamma$, which also does not exceed the semi-norm of f. $\square$

Petrushev [1988] has obtained companion Jackson and Bernstein inequalities for free knot spline approximation in L_p, $0 < p < \infty$, using the Besov spaces

$$B^\alpha := B^\alpha_\gamma(L_\gamma), \qquad \gamma := (\alpha + 1/p)^{-1}, \tag{8.7}$$

with the semi-norm given by

$$|f|_{B^\alpha} = \left(\int_0^\infty \left[t^{-\alpha}\omega_r(f,t)_\gamma\right]^\gamma \frac{dt}{t}\right)^{1/\gamma}. \tag{8.8}$$

Theorem 8.2 (Petrushev). *Let $0 < p < \infty$, $r = 1, 2, \dots$, and $0 < \alpha < r$. For the approximation by free knot splines of order r, we have*

$$\begin{cases} \text{(i)} \quad \sigma_{n,r}(f)_p \le Cn^{-\alpha}|f|_{B^\alpha} \le Cn^{-\alpha}\|f\|_{B^\alpha}, & n = 1, 2, \dots \\ \text{(ii)} \quad \|S\|_{B^\alpha} \le Cn^\alpha \|S\|_p, \quad S \in \Sigma_{n,r}, & n = 1, 2, \dots. \end{cases} \tag{8.9}$$

Proof. (i). Let $M^\gamma := \int_0^1 t^{-\alpha\gamma-1}w_r(f,t)_\gamma^\gamma\, dt$, where w_r is the averaged modulus of smoothness (5.15). From (5.17), and Theorem 10.1 we have that M is equivalent to $|f|_{B^\alpha}$ with constants of equivalency depending only on r and γ. If we let $g(x,s,t) := t^{-\alpha\gamma-2}\chi_{[0,t]}(s)\chi_{A_{rs}}(x)|\Delta_s^r(f,x)|^\gamma$, with $A_{rs} := [0, 1-rs]$, $s > 0$, then by the definition of w_r,

$$M^\gamma = \int_0^1\int_0^\infty\int_0^1 g(x,s,t)\, dxdsdt = \int_0^1 G(x)\, dx \tag{8.10}$$

where $G(x) := \int_0^\infty\int_0^\infty g(x,s,t)\, dsdt$. We choose intervals I_j, $j = 1, \dots, n$, with pairwise disjoint interiors so that $\bigcup_j I_j = A$ and $\int_{I_j} G(x)\, dx = M^\gamma/n$. Then, using once again (5.17), we have

$$|f|^\gamma_{B^\alpha(I_j)} \le C\int_{I_j} G(x)\, dx = CM^\gamma/n, \qquad j = 1, \dots, n.$$

From Theorem 8.1, there are polynomials P_j of degree $< r$ such that for all j,

$$\|f - P_j\|_p(I_j) \le C|f|_{B^\alpha(I_j)} \le CMn^{-1/\gamma}. \tag{8.11}$$

If we define $S := P_j$ on I_j, then S is in Σ_{nr} and from (8.11),

$$\sigma_{nr,r}(f)_p^p \le \sum_{j=1}^n \|f - P_j\|_p^p(I_j) \le CM^p n^{-p/\gamma} n = CM^p n^{-\alpha p}.$$

Thus, (i) follows from the monotonicity of the $\sigma_{n,r}$.
(ii). If $S \in \Sigma_{n,r}$, there exist intervals I_j, $j = 1, \dots, n$, with pairwise disjoint interiors and with $A = \cup I_j$ and polynomials P_j, $j = 1, \dots, n$, of degree $< r$ such that

$$S = \sum_{j=1}^n \phi_j; \quad \phi_j := P_j\chi_{I_j}. \tag{8.12}$$

To estimate $\omega_r(S,t)_\gamma$, we fix an $h > 0$. For given x, the points $x + kh$, $k = 0, \dots, r$ are inside at most $r + 1$ of the intervals I_j. All other I_j contain no points $x + kh$ and have $\Delta_h^r(\phi_j, x) = 0$. This implies that

$$|\Delta_h^r(S,x)|^\gamma \le C\sum_{j=1}^n |\Delta_h^r(\phi_j,x)|^\gamma. \tag{8.13}$$

Let $\Gamma_j = \{x : \Delta_h^r(\phi_j, x) \ne 0\}$. This set is contained in the union $\bigcup_{k=0}^r (I_j - kh)$, and it does not contain points $x \in I_j$ at a distance $\ge rh$ from the endpoints of I_i, because then $\Delta_h^r(P_j, x) = 0$. Hence $|\Gamma_j| \le C\min(h, |I_j|)$. From (8.13) and (2.14) of Chapter 4 we have

$$\begin{aligned} \int_{A_{rh}} |\Delta_h^r(S,x)|^\gamma\, dx &\le C\sum_{j=1}^n \|\phi_j\|_\infty^\gamma(I_j)|\Gamma_j| \\ &\le C\sum_{j=1}^n \|\phi_j\|_p^\gamma(I_j)|I_j|^{-\gamma/p}\min(h, |I_j|). \end{aligned} \tag{8.14}$$

It follows that $\omega_r(S,t)_\gamma^\gamma$ is majorized by the right side of (8.14) with h replaced by t. Therefore, using Hölder's inequality at the last step, we get

$$\begin{aligned} |S|^\gamma_{B^\alpha} &\le C\sum_{j=1}^n \|\phi_j\|_p^\gamma(I_j)|I_j|^{-\gamma/p}\int_0^\infty t^{-\alpha\gamma-1}\min(t, |I_j|)\, dt \\ &\le C\sum_{j=1}^n \|\phi_j\|_p^\gamma(I_j) \le Cn^{1-\gamma/p}\|S\|_p^{\gamma/p}. \end{aligned} \tag{8.15}$$

Since $1 - \gamma/p = \alpha\gamma$, and $\|S\|_\gamma \le \|S\|_p$, we have the desired result. □

For $p = \infty$, part (ii) of Theorem 8.2 is not valid. Indeed, then $\gamma = 1/\alpha$ and the characteristic function of an interval inside A does not belong to B^α.

We are now in a position to apply the general theory of §5 and of §9 of Chapter 7. We take

$$X = L_p, \quad Y = B^\alpha, \quad \Phi_n = \Sigma_{n,r}.$$

Inequalities (i), (ii) of (8.9) become the Jackson and the Bernstein inequalities (5.4), (5.5) (with r replaced by α) of Chapter 7. Since (5.2) of Chapter 7 holds for $\Sigma_{n,r}$, Theorem 9.1 of that chapter yields

Theorem 8.3 (Petrushev [1988]). *For free knot spline approximation of order r in L_p, $0 < p < \infty$, we have for all $0 < \alpha < \beta < r$ and $0 < q \le \infty$*

$$A_q^\alpha(L_p) = (L_p, B^\beta)_{\alpha/\beta, q}. \tag{8.16}$$

In some cases, we can explicitly determine the interpolation spaces in Theorem 8.3. We first observe that the following embeddings hold

$$A^\alpha_\gamma(L_p) \subset B^\alpha \subset A^\alpha_\infty(L_p). \tag{8.17}$$

Indeed, the Jackson inequality (i) of Theorem 8.2 is the right embedding of (8.17). To prove the left embedding, we recall that $|\cdot|^\gamma_{B^\alpha}$, $\gamma = \gamma(\alpha)$, has an equivalent semi-norm (8.10) which is subadditive. Now let $f \in B^\alpha$. From the embedding (8.3), it follows that $f \in L_p$. We take $S_n \in \Sigma_{2^n,r}$ to satisfy $\|f - S_n\|_p \le 2\sigma_{2^n,r}(f)_p$, $n = 0, 1, \ldots$. Then, if S_{-1} is the best approximation to f from $\mathcal{P}_r$, and $S_{-2} = 0$, we have $f = \sum_{k=1}^\infty (S_k - S_{k-1})$ with convergence in L_p. Since $S_0 \in \mathcal{P}_r$, $|S_0|_{B^\alpha} = 0$ and the inverse inequality (ii) of Theorem 8.2 gives

$$\begin{aligned} |f|^\gamma_{B^\alpha} &\le \sum_{n=0}^\infty |S_n - S_{n-1}|^\gamma_{B^\alpha} \le C \sum_{n=1}^\infty 2^{n\alpha\gamma} \|S_n - S_{n-1}\|^\gamma_p \\ &\le C \sum_{n=0}^\infty 2^{n\alpha\gamma} \sigma_n(f)^\gamma_p = C|f|_{A^\alpha_\gamma(L_p)} \end{aligned}$$

which gives the left embedding of (8.17). □

Theorem 8.4 (DeVore-Popov [1988$_1$]). *For free knot spline approximation in L_p of order r, with $\alpha < r$, we have*

$$A^\alpha_\gamma(L_p) = B^\alpha, \qquad \gamma = (\alpha + 1/p)^{-1}. \tag{8.18}$$

That is,

$$f \in B^\alpha \iff \sum_{n=0}^\infty [2^{n\alpha}\sigma_n(f)_p]^\gamma < \infty. \tag{8.19}$$

Proof. From Theorem 7.4 of Chapter 7, the spaces $A^\alpha_q(L_p)$ are an interpolation family. Therefore, if $0 < \beta_0, \beta_1 < r$, $0 < q, q_0, q_1 \le \infty$ and $0 < \theta < 1$, we have

$$(A^{\beta_0}_{q_0}(L_p), A^{\beta_1}_{q_1}(L_p))_{\theta,q} = A^\lambda_q(L_p), \quad \lambda := (1-\theta)\beta_0 + \theta\beta_1.$$

Therefore, the reiteration Theorem 7.3 of Chapter 6 and the embeddings (8.17) give

$$\begin{aligned} A^\alpha_q(L_p) &= (A^{\beta_0}_{\gamma(\beta_0)}(L_p), A^{\beta_1}_{\gamma(\beta_1)}(L_p))_{\theta,q} \subset (B^{\beta_0}, B^{\beta_1})_{\theta,q} \subset (A^{\beta_0}_\infty(L_p), A^{\beta_1}_\infty(L_p))_{\theta,q} \\ &= A^\alpha_q(L_p). \end{aligned}$$

We now take $\beta_0 < \alpha < \beta_1$ and choose $0 < \theta < 1$ so that $\alpha = (1-\theta)\beta_0 + \theta\beta_1$. From the definition of γ, it follows that $1/\gamma(\alpha) = (1-\theta)/\gamma(\beta_0) + \theta/\gamma(\beta_1)$. Therefore, if we take $q = \gamma(\alpha)$, Theorem 7.3 says

$$(B^{\beta_0}, B^{\beta_1})_{\theta,q} = B^\alpha$$

and (8.18) follows. □

As a corollary, we obtain an interpolation result for Besov spaces.

Corollary 8.5. *For $0 < p < \infty$, $0 < \alpha < \beta$ and $\gamma = \gamma(\alpha)$ of (8.7), we have*

$$(L_p, B^\beta)_{\alpha/\beta,\gamma} = B^\alpha \tag{8.20}$$

Proof. This follows from (8.16) and (8.18). □

§ 9. Problems

9.1. If $f \in C(A)$, $A := [0,1]$, has bounded p-variation, $0 < p \le 1$, i.e. $\sup \sum_{i=1}^m |f(x_i) - f(x_{i-1})|^p < \infty$, with the supremum taken over all $0 =: x_0 < x_1 < \cdots < x_m := 1$, then for approximation by piecewise constants with n free knots: $\sigma_n(f)_\infty \le Cn^{-1/p}$.

9.2. For equally spaced knots $T_n := (j/n)_1^{n-1}$ in $A = [0,1]$, if $f \in W^r_p(A)$, $1 \le p \le \infty$, there is an $S_n \in \mathcal{S}_r(T_n, A)$ such that $\|f^{(k)} - S_n^{(k)}\|_p \le C_r \|f^{(r)}\|_p n^{-r+k}$, $0 \le k < r$.

9.3. (DeVore-Richards [1972]) Let T_n, $n = 1, 2, \ldots$, be knot sequences with mesh lengths δ_n monotonically decreasing to zero and satisfying $\delta_n/\delta_{n+1} \le M$, $n = 1, 2, \ldots$. If for each point $x \in A$, $\operatorname{dist}(x, T_m) \ge C\delta_n$ for some $m = m(x) \ge n$, then $E_r(f, T_n)_\infty = o(\delta_n^r)$ implies that f is a polynomial of degree $< r$.

9.4. There is a function $f \in W^1_1$ which cannot be adaptively approximated with error $\mathcal{O}(n^{-1})$ in the uniform norm by piecewise constants with n values.

9.5. For each $f \in W^r_1(A)$, $A := [0,1]$, $r = 2, 3, \ldots$, there is a knot sequence $T_n = (t_i)_1^n$ with locally bounded mesh ratios: $|t_{i+1} - t_i| \le C|t_{j+1} - t_j|$, $j = i-1, i+1$ and a spline $S \in \mathcal{S}_r(T_n, A)$ for which $\|f - S\|_\infty(A) \le C_r n^{-r}$.

9.6. For each $0 < p < 1$, there is no constant $C > 0$ for which $\omega(f,t)_p \le C\|f'\|_p t$ would held for all $f \in C^1[0,1]$, $t > 0$.

9.7. For each $f \in W^r_1(A)$, $A = [0,1]$, and $r = 1, 2, \ldots$, there is a spline $S_n \in \Sigma_{n,r} \cap C^{r-1}$ with the properties $\|f - S_n\|_\infty \le C_r n^{-r}$ and $\operatorname{Var}(S_n^{(r-1)}) \le C_r \|f^{(r)}\|_1$.

§ 10. Notes

10.1. The first results on adaptive approximation are due to Birman and Solomyak [1967]. Direct theorems for adaptive approximation require only slightly more smoothness than for free knot splines. For example (compare with Kahane's Theorem 4.2), DeVore [1987] has shown that if $f' \in L \log L$, then f can be approximated in the uniform norm with order $\mathcal{O}(n^{-1})$ adaptively by piecewise constants with at most n pieces. Error estimates for adaptive approximation to functions with singularities have been given by de Boor and Rice [1979]. There are no known effective inverse theorems for adaptive approximation.

10.2. There is an analogue of free knot spline approximation in which not only the knots but also the degree of the polynomial pieces of a spline are allowed to vary. DeVore and Scherer [1980] have introduced the class of piecewise polynomials $\tilde{\Sigma}_n$ which have m polynomial pieces (on $A = [0,1]$) of degree n_i, $i = 1, \ldots, m$, and $\sum_{i=1}^m n_i \le n$. They show that for $\alpha > 0$, the functions x^α satisfy: $\tilde{\sigma}_n(x^\alpha) \sim \exp(-c\sqrt{\alpha n})$, $c := 2\ln\sqrt{2} - 1$.

10.3. Barrow, Chui, Smith, and Ward [1978] have shown that in some cases the L_2 approximation is unique from the nonlinear manifold Σ_n of linear splines with $n+1$ knots $0 = t_0 < t_1 < \cdots < t_n = 1.$. They prove that if $f \in C^2[0,1]$, if $f'' > 0$, and if $\log f''$ is concave on $[0,1]$, then for any N, f has a unique best L_2 approximation from Σ_n. It is not sufficient to know that $f \in C^\infty[0,1]$ and $f'' > 0$. However for a function $f \in C^5[0,1]$ with $f'' > 0$, the best L_2 approximation from Σ_n is unique for all sufficiently large n.

Chapter 13. Spline Interpolation and Projections onto Spline Spaces

§ 1. Introduction. Lagrange Interpolation by Splines

One of the simplest and most important ways to approximate an element f of the Banach space X by elements from a subspace Y is by means of a projection P from X onto Y. Such a projection is a bounded linear operator from X onto Y which is the identity on Y. From Lebesgue's Proposition 4.1 of Chapter 2, we see that Pf is good for approximation to f if the norm of P is small. Thus in Chapter 9 we have computed or estimated norms of several projections. Here we do this for projections onto spline spaces.

We have shown in Chapter 5, Theorem 4.4, that the de Boor-Fix quasi-interpolants are projections onto the Schoenberg spaces $\mathcal{S}_r(T, I)$ whose L_p norms depend only on the degree of the splines (and not on their knot spacing). In this chapter, we discuss other more natural projections onto spline spaces. In §§1–2, we give bounds for the norms of the Lagrange spline interpolation operators and study their dependence on the choice of knots. The estimates for interpolation by cubic splines in §3 are particularly simple but important. The least squares approximation, in other words the orthogonal projection onto $\mathcal{S}_r(T)$, is the main subject of §§4–5. In these sections, as well as later, properties of finite or infinite matrices play an important role. In §6, we have the main theorems of Schoenberg on cardinal spline interpolation, when the knots are the integers; §7 is a generalization thereof. The last §§8–9 deal with shape preserving approximation (by means of interpolation) by low degree splines. Several algorithms that we describe are useful for numerical curve fitting.

In this and the next two sections, $\mathcal{S}_r(T, I)$, $I = [0,1]$, will be a *Schoenberg spline space with simple knots* $0 < t_1 < \cdots < t_n < 1$; we put $t_0 := 0$, $t_{n+1} := 1$, and, in case that B-splines will be used, supplement this to a sequence $T := (t_i)_{-r+1}^{n+r}$ with auxiliary knots defined by $t_i := 0$, $i < 0$, $t_i := 1$, $i > n_0$.

We shall discuss projections onto $\mathcal{S}_r(T, I)$ given by Lagrange interpolation. Let $X : 0 \le x_{-r+1} < \cdots < x_n \le 1$ be interpolation points and $(y_i)_{-r+1}^{n}$ the corresponding data. By the Schoenberg-Whitney theorem (Theorem 9.1 of Chapter 5), there is a unique spline $S \in \mathcal{S}_r(T, I)$ satisfying

$$S(x_i) = y_i\,, \qquad i = -r+1, \ldots, n\,, \tag{1.1}$$

if and only if the interpolation points interlace with the knots:

$$x_i \in (t_i, t_{i+r}), \qquad i = -r+1, \dots, n. \tag{1.2}$$

If the $y_i = f(x_i)$ are given by functions $f \in C(I)$, we obtain an interpolation operator $L := L(X,T)$ which maps functions $f \in C(I)$ onto splines $S \in \mathcal{S}_r(T,I)$. Obviously, this is a projection onto $\mathcal{S}_r(T,I)$ which is bounded in the L_∞ norm. We would like to know how this norm, $\|L(X,T)\|$, depends on X and T. In particular, for which choices of X, T will this norm have an upper bound which depends only on the order r of the splines?

The only completely trivial case is when $r = 1$. Then $\|L(X,T)\| = 1$ for all T and X satisfying (1.2). Indeed, in this case, $L(f) = \sum_0^n f(x_i)\chi_i$ where χ_i is the characteristic function of (t_i, t_{i+1}), $i = 0, \dots, n$. More illuminating is the case $r = 2$ of broken line interpolation. It can happen that two interpolation points x_j, x_{j+1} are in the same interval (t_i, t_{i+1}). When this is the case, $\|L(X,T)\|$ will become arbitrarily large as $x_j \to x_{j+1}$ if T is fixed. For example, if f is a norm one function which is 1 at x_{j+1} and 0 at all other interpolation points, then $L(f)$ will be linear on (t_i, t_{i+1}) with slope $a := (x_{j+1} - x_j)^{-1}$. The value of $|L(f)(t)|$ will be at least $(a/2)(t_{i+1} - t_i)$ at one of the points t_i, t_{i+1}, so that $\|L(X,T)\| \to \infty$ as $x_j \to x_{j+1}$. This phenomenon persists for all $r \geq 2$: whenever the knots are held fixed and the interpolation points are allowed to coalesce $\|L(X,T)\|$ will become unbounded (see the discussion after Theorem 1.2).

There is an important connection between the projection L and the *collocation matrix* $C^* := C^*(X,T)$ associated with the interpolation problem (1.1). It is a square matrix defined in (9.7) of Chapter 5 by means of

$$C^* := C^*(X,T) := (N_j(x_i)), \quad i,j = -r+1, \dots, n. \tag{1.3}$$

Here $N_j(x) := N(x; t_j, \dots, t_{j+r})$, $j = -r+1, \dots, n$, are the B-splines of order r for the knot sequence T.

Let $L(f) = \sum_{j=-r+1}^n c_j N_j$ with $c_j := c_j(f)$, be the B-spline representation for the operator L. With $c = (c_j)$, one has

$$(L(f)(x_i)) = C^* c. \tag{1.4}$$

An alternative formulation of the Schoenberg-Whitney theorem is that the matrix C^* is invertible if and only if (1.2) is satisfied. In this case, we have a one-to-one correspondence between the splines $L(f)$, the coefficients c and the values $(f(x_i))$ of f.

The theorem of de Boor that follows shows that $\|L\|$ is comparable in size to $\|C^{*-1}\|_\infty$. To prove this, we shall use the *condition number* D_r of the B-spline basis. Let $1 \leq p \leq \infty$ and let

$$d_j := \frac{t_{j+r} - t_j}{r}. \tag{1.5}$$

From (4.9) of Chapter 5, for any sequence $c := (c_j)_{-r+1}^n$:

$$D_r \|c\|_p \leq \Big\| \sum_{-r<j\leq n} c_j d_j^{-\frac{1}{p}} N_j \Big\|_p (I) \leq \|c\|_p. \tag{1.6}$$

For a square matrix $B := (b_{i,j})$, $1 \leq i, j \leq n$, to each $x = (x_i)_1^n \in l_p^{(n)}$, there corresponds a $y = (y_i)_1^n$ given by $y = Bx$. The norm of the operator B as a mapping from $l_p^{(n)}$ into itself is given by the formulas

$$\begin{aligned} \|B\|_p : &= \sup\left\{\frac{\|Bx\|_p}{\|x\|_p} : x \neq 0\right\} \\ &= \sup\{y^t Bx : \|x\|_p, \|y\|_{p'} \leq 1\}, \quad \frac{1}{p} + \frac{1}{p'} = 1. \end{aligned} \tag{1.7}$$

The same remarks apply to bi-infinite matrices $B = (b_{ij})$ provided that $y_i = \sum b_{ij} x_j$, $x \in l_p$, converges for all i, and $y \in l_p$. It is easy to compute these norms when $p = 1$ or $p = \infty$:

$$\|B\|_\infty = \max_i \sum_j |b_{i,j}|, \quad \|B\|_1 = \max_j \sum_i |b_{i,j}|. \tag{1.8}$$

For other values of p, there is no simple expression for $\|B\|_p$ in terms of the entries $b_{i,j}$. For example, it is known that $\|B\|_2$ is the spectral norm of B which is the square root of the absolute value of the largest eigenvalue of $B^t B$ (see for example Strang [B-1976, p. 272]).

For two matrices A, B, $\|AB\|_p \leq \|A\|_p \|B\|_p$. In particular, if A has an inverse $A^{-1} : AA^{-1} = I$, then

$$\kappa_p(A) := \|A\|_p \|A^{-1}\|_p \geq 1. \tag{1.9}$$

This *condition number* measures how far A is removed from the identity matrix.

Theorem 1.1 (de Boor [1975]). *If X, T satisfy the interlacing condition* (1.2), *then with the constant D_r of* (1.6), *we have for the spline interpolation operator $L := L(X,T)$*

$$D_r \|C^{*-1}\|_\infty \leq \|L\|_\infty \leq \|C^{*-1}\|_\infty. \tag{1.10}$$

Proof. For $a := (a_i)_{-r+1}^n$, we have

$$\|C^{*-1}\|_\infty = \sup\Big\{\frac{\|a\|_\infty}{\|C^* a\|_\infty} : a \neq 0\Big\}.$$

Let X, T be given. By definition,

$$\|L\| = \sup_{f \neq 0} \frac{\|L(f)\|_\infty}{\|f\|_\infty}. \tag{1.11}$$

The spline $L(f)$ is defined by the values $y_i = f(x_i)$. Also if the y_i are given there is a function f with $f(x_i) = y_i$ and $\|f\|_\infty = \max_i |f(x_i)|$. Therefore, in (1.11), we can replace the denominator by $\max_i |L(f)(x_i)|$. If $L(f) = \sum_{-r+1}^n c_j N_j$, we can use (1.4) and obtain

$$\|L\| = \sup \frac{\left\| \sum_{-r+1}^{n} c_j N_j \right\|_\infty}{\|C^* c\|_\infty} .$$

According to (1.6) the numerator of the last expression lies between $D_r\|c\|_\infty$ and $\|c\|_\infty$; hence (1.10) follows. □

We can use Theorem 1.1 to give a lower bound for $\|L\|$ in terms of the spacing of the interpolation points and the knots. We let

$$\delta_i := \min\{t_{j+r-1} - t_j : (t_j, t_{j+r-1}) \cap (x_i, x_{i+1}) \neq \emptyset\} . \tag{1.12}$$

Theorem 1.2 (de Boor [1975]). *For the interpolation operator L of order $r > 1$, we have, for all X, T,*

$$\|L\| \geq \frac{D_r}{(r-1)} \max_{-r<i\leq n} \frac{\delta_i}{x_{i+1} - x_i} . \tag{1.13}$$

Proof. We fix i with $-r < i \leq n$ and let $f \in C(I)$ be a norm one function which satisfies: $f(x_i) = 1$, $f(x_j) = -1$, $j = -r+1, \ldots, n$, $j \neq i$. We can express $S := L(f)$ in its B-spline series $S = \sum_{-r+1}^{n} c_j N_j$. From (3.11) of Chapter 5, we have the following representation for S' with $c_{-r} := c_{n+1} := 0$ and $N_{j,r-1}$ the B-splines of order $r-1$ for the knot sequence T:

$$S' = (r-1) \sum_{-r+1}^{n+1} \frac{c_j - c_{j-1}}{t_{j+r-1} - t_j} N_{j,r-1} .$$

Since $\sum N_{j,r-1} = 1$ on I, it follows from (1.6), and the definition of δ_i that for $x \in (x_i, x_{i+1})$

$$\frac{1}{r-1}|S'(x)| \leq \max \frac{|c_j - c_{j-1}|}{t_{j+r-1} - t_j} \leq \frac{2\|c\|_\infty}{\delta_i} \leq \frac{2D_r^{-1}\|S\|_\infty}{\delta_i}$$

where the maximum is taken over all j such that $N_{j,r-1}(x) \neq 0$, and therefore $(t_j, t_{j+r-1}) \cap (x_i, x_{i+1}) \neq \emptyset$. Hence,

$$2 = |f(x_{i+1}) - f(x_i)| \leq \int_{x_i}^{x_{i+1}} |S'(t)| dt \leq 2(r-1) D_r^{-1} (x_{i+1} - x_i) \|S\|_\infty \delta_i^{-1} .$$

Since i is arbitrary we have (1.13). □

The lower bound (1.13) confirms a remark we made earlier; namely, if the knots T are fixed, then $\|L(X,T)\|$ will become unbounded if two of the interpolation points move towards coalescing. Even more disturbing is the fact that for given interpolation points X, it may not be possible to choose a constant C_r so that $\|L(X,T)\| \leq C_r$ for all T. For such an example, let $r > 2$. Then, given any $M \geq 1$, we choose a, b so that $0 < 2Ma < b$. If the interpolation points X satisfy $|x_{2i+1} - x_{2i}| \leq a$ and $|x_{2i-1} - x_{2i}| \geq b$ for all i, then the maximum in (1.13) is $\geq M$, no matter how the knots T are chosen. Indeed, if this were not the case, we would have $\delta_{2i} < Ma$ for all i. This means that for each i there are at least r knots within a distance Ma of x_{2i} and that none of these knots is within a distance Ma of any other x_{2j}, $j \neq i$. But then the number of knots would be of the order $r/2$ times the number of interpolation points. This is impossible when the number of interpolation points is large.

We meet with more success if instead we are given the knots T and we wish to select interpolation points X (depending on T) in such a way that $\|L(X,T)\| \leq C_r D_r$. According to (1.13), we must have

$$\max_i \frac{\delta_i}{x_{i+1} - x_i} \leq C_r . \tag{1.14}$$

It is easy to see that there are always points X for which (1.14) and (1.2) are satisfied. Indeed, this is the case for the "knot averages" of T:

$$x_i := \frac{t_{i+1} + \cdots + t_{i+r-1}}{r-1}, \quad i = -r+1, \ldots, n . \tag{1.15}$$

For this choice we have (1.2), moreover, $x_{i+1} - x_i = \frac{t_{i+r} - t_{i+1}}{r-1}$ and (1.14) is satisfied for $C_r = r-1$. We shall discuss this and other selections of interpolation points in the section that follows.

§ 2. Selection of Interpolation Points

Marsden [1974] has shown that for quadratic splines on $I := [0,1]$, interpolation at the knot averages (1.15) yields a projection whose norm is bounded independently of T.

Theorem 2.1 (Marsden). *If $r = 3$ and T is arbitrary, then for the knot averages $x_i := \frac{1}{2}(t_{i+1} + t_{i+2})$ of (1.15), we have*

$$\|L(X,T)\| \leq 18 . \tag{2.1}$$

To prove this theorem we will exploit (1.10) and the total positivity of the collocation matrix C^* (see Theorem 11.1 of Chapter 5). We have the following useful lemma of de Boor [1968].

Lemma 2.2. *Let A be an $m \times m$ totally positive invertible matrix. If there is a vector $u := (u_i)_1^m$ satisfying $\|u\|_\infty = 1$ and $Au = v$ with $(-1)^i v_i \geq \delta > 0$, $i = 1, \ldots, m$, then*

$$\|A^{-1}\|_\infty \leq \frac{1}{\delta} . \tag{2.2}$$

Proof. Let $B := A^{-1} =: (b_{i,j})$. The total positivity of A and Cramer's rule show that B has the checkerboard property: $(-1)^{i+j}b_{i,j} \geq 0$, $i,j = 1,\ldots,m$. The norm $\|B\|_\infty$ is given by (1.8). To estimate one of the sums, we use the vector $v := C^*u$ in our assumptions. We have

$$\sum_{j=1}^m |b_{i,j}| = \sum_{j=1}^m (-1)^{i+j} b_{i,j} \leq \sum_{j=1}^m (-1)^{i+j} b_{i,j}(-1)^j \frac{v_j}{\delta} = (-1)^i \frac{u_i}{\delta} \leq \frac{1}{\delta}. \qquad \square$$

Proof of Theorem 2.1. We shall find a vector $c := (c_j)_{-r+1}^n$ with $\|c\|_\infty = 2$ such that $S(x) := \sum c_j N_j$ satisfies

$$(2.3) \qquad (-1)^i S(x_i) \geq \frac{1}{9}, \quad i = -r+1,\ldots,n.$$

Since for the collocation matrix C^* of (1.3), $S(x_i) = (C^*c)_i$, $i = -r+1,\ldots,n$, (2.1) will follow from (1.10) and Lemma 2.2.

To find c, we use the de Boor-Fix formula (3.9) of Chapter 5. If we take $\xi_i := x_i$ in this formula, we have

$$(2.4) \qquad c_i = S(x_i) - (\Delta t_{i+1})^2 S'' \frac{(x_i)}{8}, \quad i = -r+1,\ldots,n.$$

We can compute S'' from (3.11) of Chapter 5:

$$(2.5) \qquad S''(x) = 2\sum_j \left(\frac{\Delta c_{j-1}}{t_{j+2}-t_j} - \frac{\Delta c_{j-2}}{t_{j+1}-t_{j-1}}\right) \frac{1}{\Delta t_j} N_{j,1}(x).$$

For $x = x_i = \frac{t_{i+1}+t_{i+2}}{2}$, only the term $j = i+1$ in (2.5) is non-zero; hence from (2.4), we have for $i = -r+1,\ldots,n$,

$$(2.6) \qquad S(x_i) = c_i + \frac{b_i}{4}; \quad b_i := \frac{\Delta t_{i+1}\Delta c_i}{t_{i+3}-t_{i+1}} - \frac{\Delta t_{i+1}\Delta c_{i-1}}{t_{i+2}-t_i}.$$

In computing the Δc_i, we take $c_{-r} := 0$ and $c_{n+1} := 0$.

We can now define our sequence c. For $i = -r+1,\ldots,n$, we let

$$c_i := (-1)^i \begin{cases} 2, & \text{if } \Delta t_i \leq 8\Delta t_{i+1} \\ 1, & \text{otherwise,} \end{cases}$$

and show that (2.3) holds for this choice. If $|c_i| = 1$, then $\Delta t_{i+1} \leq \frac{(t_{i+2}-t_i)}{8}$ and so $|b_i| \leq 3+\frac{3}{8}$. Using this in (2.6), we obtain $(-1)^i S(x_i) \geq \frac{5}{32} \geq \frac{1}{9}$. On the other hand, if $|c_i| = 2$ and $|c_{i+1}| = 1$, then $|b_i| \leq 7$; while if $|c_i| = |c_{i+1}| = 2$, then $\Delta t_{i+1} \leq 8\Delta t_{i+2}$ and so $\Delta t_{i+1} \leq 8\frac{(t_{i+3}-t_{i+1})}{9}$ and $|b_i| \leq 4+\frac{32}{9}$. These two estimates for b_i give $(-i)^i S(x_i) \geq \frac{1}{9}$. $\square$

There are explicit choices of interpolation points which result in uniformly bounded projections for other values of r (for example the knot averages (1.15) work when $r = 4$ (de Boor [1975]). However, there are no explicit formulas for interpolation points that are known to work for all r. On the other hand, the following theorem of Demko [1985] (also proved independently by Mørken [1984]) shows that such points exist.

Theorem 2.3. *For each $r = 1,2,\ldots$ and each knot sequence $T := (t_i)$, there are interpolation points X for which $\|L(X,T)\| \leq D_r^{-1}$, where D_r is the constant in* (1.6).

Proof. Let N_j, $j = -r+1,\ldots,n$ be the B-splines for T. According to Theorem 10.2 of Chapter 5, the splines N_j, $j = -r+1,\ldots,n-1$, are a weak Haar system on I. The Jones-Karlovitz theorem (Theorem 8.6 of Chapter 10) asserts that there is a best uniform approximation S_0^* to N_n from the span of the N_j, $j = -r+1,\ldots,n-1$, such that $S^* := N_n - S_0^*$ has at least $n+r$ alternation points $X : 0 \leq x_1 < \cdots < x_{n+r} \leq 1$. That is,

$$(2.7) \qquad \eta(-1)^i S^*(x_i) = \|S^*\|_\infty, \quad i = 1,\ldots,n+r,$$

for some $\eta = \pm 1$. We let $S := \eta S^*$ and write $S = \sum_{-r+1}^n c_j N_j$. Then, from (1.6), we have

$$\|c\|_\infty \leq D_r^{-1}\|L\|_\infty.$$

Since for the collocation matrix of (1.3), $(C^*c)_i = S(x_i)$, $i = 1,\ldots,n+r$, the theorem follows from (2.7), Lemma 2.2, and Theorem 1.1. $\square$

The existence of the spline S^* can also be derived easily from Borsuk's theorem as was done in the original argument of Demko.

§3. Cubic Spline Interpolation

In most applications, the interpolation points are given in advance. As we have seen in the previous section, it may not be possible to choose knots so that the interpolation operator L would be bounded on $C(I)$. However, $L(f)$ is still usually a good approximation to f provided f has sufficient smoothness. We shall give some examples of this for cubic splines (that is, splines of order $r = 4$). This case is particularly useful in numerical applications and therefore we take some care to give realistic constants in our estimations.

For given interpolation points $X : 0 = x_0 < x_1 < \cdots < x_n = 1$, we wish to find a cubic spline S which interpolates the values y_i at the points x_i:

$$(3.1) \qquad S(x_i) = y_i, \quad i = 0,1,\ldots,n.$$

We usually assume that $y_i := f(x_i)$ for some function f.

We have yet to specify the knots of S. The most widely used cubic spline interpolants have their knots at the points x_i. We shall therefore begin by considering piecewise polynomials S with breakpoints at the x_i which satisfy the interpolation conditions (3.1). We assume in addition that $S \in C^1(I)$ and therefore S is determined by the values $y_i := S(x_i)$ and the derivative values $s_i := S'(x_i)$, $i = 0,\ldots,n$. We let

$$(3.2) \qquad \phi(x) := 1 + x^2(2x-3), \quad \psi(x) := x(1-x)^2$$

be the Hermite cubic polynomials. They satisfy $\phi(0) = \psi'(0) = 1$ and all other values of ϕ, ψ, ϕ', ψ' at 0 and 1 are zero.

On the interval $I_i := [x_i, x_{i+1}]$ of length $\Delta x_i := x_{i+1} - x_i$, we have

$$S(x) = P_i(x) := y_i \phi\left(\frac{x-x_i}{\Delta x_i}\right) + s_i \Delta x_i \psi\left(\frac{x-x_i}{\Delta x_i}\right) + y_{i+1}\phi\left(\frac{x_{i+1}-x_i}{\Delta x_i}\right)$$

(3.3)

$$- s_{i+1}\Delta x_i \psi\left(\frac{x_{i+1}-x}{\Delta x_i}\right), \qquad i = 0, \dots, n-1.$$

The values y_i are given but we have freedom in selecting the slopes s_i. If we are interpolating a function f and the values $f'(x_i)$ are available to us, we may simply take $s_i := f'(x_i)$. The resulting spline $H := H(f)$ is the *Hermite cubic interpolant* to f. It is in $C^1(I)$ and also in the Schoenberg space $\mathcal{S}_4(T, I)$ where the basic knots consist of the x_i, $i = 1, \dots, n-1$, each repeated twice.

We give next a simple estimate for the error in Hermite cubic interpolation. From (3.3), we have $H(f,x) = Q_i(f,x)$, $x \in [x_i, x_{i+1}]$ where $Q_i := Q_i(f, \cdot)$, $i = 0, \dots, n-1$, are the *Hermite cubic polynomials:*

$$Q_i(f,x) := f(x_i)\phi\left(\frac{x-x_i}{\Delta x_i}\right) + f'(x_i)\Delta x_i \psi\left(\frac{x-x_i}{\Delta x_i}\right)$$

(3.4)

$$+ f(x_{i+1})\phi\left(\frac{x_{i+1}-x}{\Delta x_i}\right) - f'(x_{i+1})\Delta x_i \psi\left(\frac{x_{i+1}-x}{\Delta x_i}\right).$$

The polynomial Q_i is completely determined by the values of f on $[x_i, x_{i+1}]$. From Newton's formula ((7.3) of Chapter 4) for the remainder in Hermite interpolation, we have (all norms in this section are in $C(I)$ unless otherwise indicated)

(3.5)

$$\begin{aligned}|f(x) - Q_i(f,x)| &\le |[x_i x_i, x_{i+1}, x_{i+1}, x] f|\,(x-x_i)^2 (x-x_{i+1})^2 \\ &\le \left\|f^{(4)}\right\|_\infty [x_i, x_{i+1}] \frac{\Delta x_i^4}{384}, \quad x \in [x_i, x_{i+1}],\end{aligned}$$

because the divided difference does not exceed $\frac{1}{4!}\left\|f^{(4)}\right\|_\infty [x_i, x_{i+1}]$ and $(x - x_i)(x_{i+1} - x) \le (\Delta x_i)^2/4$. We can also estimate the error of approximation for less smooth functions.

Theorem 3.1. *If $f \in C^1[x_i, x_{i+1}]$, then*

(3.6) $$|f(x) - Q_i(f,x)| \le \tfrac{9}{4}\Delta x_i E_3(f', [x_i, x_{i+1}]), \quad x \in [x_i, x_{i+1}].$$

Proof. Since $\phi'(x) = \phi'(1-x)$, $x \in [0,1]$, we have for $x \in [x_i, x_{i+1}]$,

$$Q_i(f,x)' = f'(x_i)\psi'\left(\frac{x-x_i}{\Delta x_i}\right) + f'(x_{i+1})\psi'\left(\frac{x_{i+1}-x}{\Delta x_i}\right) - \frac{\Delta y_i}{\Delta x_i}\phi'\left(\frac{x-x_i}{\Delta x_i}\right).$$

Since $|\psi'| \le 1$ and $|\phi'| \le \frac{3}{2}$ on $[0,1]$, we have in the norm of $L_\infty[x_i, x_{i+1}]$,

$$\|Q_i(f)'\| \le \tfrac{7}{2}\|f'\|.$$

If P is any cubic polynomial, we obviously have $Q_i(P) = P$. Therefore,

$$\|f' - Q_i(f)'\| \le \|f' - P'\| + \|Q_i(f-P)'\| \le \tfrac{9}{2}\|f' - P'\|.$$

Taking an infimum over all cubic polynomials P, we have

$$\|f' - Q_i(f)'\| \le \tfrac{9}{2}E_3(f', [x_i, x_{i+1}]).$$

We obtain (3.6) from our last estimate by integrating from x to the endpoint x_i, x_{i+1} closest to x. □

In (3.6), we can apply any of our estimates for $E_3(f')$ given in previous chapters. For example from Whitney's theorem (Theorem 4.1 of Chapter 6), we have

(3.7) $$\|f - H(f)\|(I) \le Ch\omega_3(f', h), \quad h := \max_i \Delta x_i.$$

It is sometimes desirable to have second order smoothness for the cubic spline interpolants. This can be obtained if the slopes s_i are chosen so that $P''_{i-1}(x_i) = P''_i(x_i)$, $i = 1, \dots, n-1$. Then, (3.3) with $z_i := 3\left(\frac{\Delta x_i}{\Delta x_{i-1}}(y_i - y_{i-1}) + \frac{\Delta x_{i-1}}{\Delta x_i}(y_{i+1} - y_i)\right)$ gives the following equations for s_i:

(3.8) $$s_{i-1}\Delta x_i + s_i 2(\Delta x_{i-1} + \Delta x_i) + s_{i+1}\Delta x_{i-1} = z_i, \quad i = 1, \dots, n-1.$$

We need two additional conditions to determine the slopes s_i. These usually consist of prescribing s_0 and s_n. There are many possibilities, but we shall focus on the case of the *complete cubic spline interpolant*, which corresponds to the choice

(3.9) $$s_0 := f'(x_0); \; s_n := f'(x_n),$$

assuming that these values are available (when they are not available they are replaced by a numerical approximation).

The coefficient matrix for the system of equations (3.8), (3.9) is tridiagonal with strictly dominant diagonal and therefore easy to solve numerically. Once the values $(s_i)_0^n$ have been determined, we can compute S from (3.3). Then, S is in the Schoenberg space $\mathcal{S}_4(T, I)$, where the basic knots are the x_i, $i = 1, \dots, n-1$. We shall establish the approximation properties of S.

Lemma 3.2. *If $f \in \mathrm{Lip}\,1$ on I then the complete cubic spline interpolant S to f satisfies in the L_∞ norm on I:*

(3.10) $$\|S'\| \le \tfrac{15}{4}\|f'\|.$$

Proof. If $M' := \max_{0<i<n}|s_i|$, we have

(3.11) $$M' \le 3\|f'\|.$$

Indeed, we can assume that $\|f'\| \le M'$ and take i so that $|s_i'| = M'$. For this value of i, the left side of (3.8) is in absolute value greater than $M'(\Delta x_{i-1} + \Delta x_i)$ and since $|z_i| \le 3\|f'\|(\Delta x_{i-1} + \Delta x_i)$, inequality (3.11) follows.

We can also estimate $s_{i+\frac{1}{2}} := S'(x_{i+\frac{1}{2}})$ where $x_{i+\frac{1}{2}}$ is the center of $[x_i, x_{i+1}]$. For this purpose, we use the identity

$$g'\left(\tfrac{1}{2}\right) = \tfrac{3}{2}\left[g(1) - g(0)\right] - \tfrac{1}{4}\left(g'(0) + g'(1)\right)$$

which holds for all cubic polynomials g. The corresponding identity on $[x_i, x_{i+1}]$ together with (3.11) gives

$$\begin{aligned} |s_{i+\frac{1}{2}}| &\leq \frac{3}{2\Delta x_i}\,|y_{i+1} - y_i| + \frac{1}{4}\left(|s_i| + |s_{i+1}|\right) \\ &\leq \frac{3}{2}\|f'\| + \frac{1}{2}M' \leq 3\|f'\|\,. \end{aligned} \tag{3.12}$$

For any quadratic polynomial Q, we have

$$|Q(x)| \leq \tfrac{5}{4}\max\left\{|Q(x_i)|, |Q(x_{i+\frac{1}{2}})|, |Q(x_{i+1})|\right\}, \quad x \in [x_i, x_{i+1}]\,. \tag{3.13}$$

It is sufficient to prove this when $x_i = 0$, $x_{i+\frac{1}{2}} = 1/2$ and $x_{i+1} = 1$ and $1/2 \leq x \leq 1$. If M is the maximum in (3.13), by means of the Lagrange interpolation formula we obtain

$$|Q(x)| \leq M\left(\frac{(x-\frac{1}{2})(1-x)}{\frac{1}{2}} + \frac{x(1-x)}{\frac{1}{4}} + \frac{x(x-\frac{1}{2})}{\frac{1}{2}}\right)$$

and the polynomial in brackets is $-4x^2 + 6x - 1 \leq 5/4$, $1/2 \leq x \leq 1$. A similar estimate applies when $0 \leq x \leq 1/2$. We use (3.13) with $Q(x) = S'(x)$. Then (3.10) follows from (3.11) and (3.12). □

The following theorem of Meir and Sharma [1969] estimates the error in approximating f by its complete cubic spline interpolant. We let $\mathcal{S}_r(X, I)$, $r = 3, 4$, be the Schoenberg spaces with basic knots $x_1, \ldots, x_n$ and let $\delta := \max_{0 \leq i < n} \Delta x_i$.

Theorem 3.3. *If $f \in C^1(I)$, then its complete cubic spline interpolant $S := L(f)$ satisfies in the uniform norm on I:*

$$\left\|f^{(j)} - S^{(j)}\right\| \leq \frac{19}{4}\left(\frac{\delta}{2}\right)^{1-j} E(f', \mathcal{S}_3(X, I)), \quad j = 0, 1\,. \tag{3.14}$$

Proof. If $R \in \mathcal{S}_4(X, I)$, then $L(R)' = R'$. Hence from (3.10),

$$\|f' - L(f)'\| \leq \|f' - R'\| + \|L(f - R)'\| \leq \frac{19}{4}\|f' - R'\|\,.$$

Since R' is an arbitrary element of $\mathcal{S}_3(X, I)$, this yields (3.14) for $j = 1$. We integrate $f' - S'$ from the knot $a = x_i$ closest to x:

$$|f(x) - S(x)| \leq \left|\int_a^x (f' - S')dt\right| \leq \frac{\delta}{2}\|f' - S'\|$$

and (3.14) for $j = 0$ follows. □

§4. Orthogonal Projections onto Splines

This projection is one of the most natural ways to approximate functions by splines. As we shall see in this and the next section, it is also useful in spline interpolation.

In this section, as well as in §§5–7, the functions will be defined on the infinite interval $\mathbb{R}$, and $T := (t_i)$ will be a bi-infinite sequence of (not necessarily simple) knots with $|t_i| \to \infty$ as $|i| \to \infty$. We let $\mathcal{S}_r := \mathcal{S}_r(T, \mathbb{R})$ be the Schoenberg space for T and r.

The L_2 spline projection $P := P_T$ assigns to each $f \in L_2(\mathbb{R})$ its best L_2 approximation Pf from $\mathcal{S}_r \cap L_2(\mathbb{R})$. Then P has norm 1 as an operator from $L_2(\mathbb{R})$ into itself. But what can we say concerning the norm of P as a mapping on other L_p spaces? Especially important is the case $p = \infty$. One of the conjectures of de Boor is that the norm of P has an upper bound that does not depend on T:

$$\|P_T\|_\infty \leq C_r\,. \tag{4.1}$$

Thus far (4.1) has only been verified for small values of r or with some additional assumptions on T. For example, in Theorem 4.4 that follows, it is shown that (4.1) holds for all T which satisfy

$$\sup_{i,j} \frac{t_{i+r} - t_i}{t_{j+r} - t_j} \leq M\,, \tag{4.2}$$

where M is some fixed constant.

To describe the projection P, we shall use the B-spline representation of the elements of $\mathcal{S}_r$. Let $N_j := N(x; t_j, \ldots, t_{j+r})$ be the B-splines of order r for the knot sequence T. We have shown in Theorem 4.2 of Chapter 5 that each $S \in \mathcal{S}_r \cap L_p$, $1 \leq p \leq \infty$, has a unique representation

$$S = \sum \alpha_j d_j^{-\frac{1}{p}} N_j\,. \tag{4.3}$$

(In this and the next sections, all sequences are bi-infinite, and all integrals are over $\mathbb{R}$.) Moreover, with the constants D_r of §4, Chapter 5, if $\boldsymbol{\alpha}$ is $(\alpha_i)_{j\in\mathbb{Z}}$,

$$D_r\|\boldsymbol{\alpha}\|_p \leq \|S\|_p \leq \|\boldsymbol{\alpha}\|_p\,. \tag{4.4}$$

That is, the splines $d_j^{-1/p}N_j$, $j \in \mathbb{Z}$, are an unconditional basis for $\mathcal{S}_r \cap L_p$. For each coefficient sequence $\boldsymbol{\alpha}$, the sum (4.3) converges both in the L_p norm and also uniformly on compact sets because at a point $x \in \mathbb{R}$ at most r terms of this sum are non-zero.

The projection P is completely described by orthogonality conditions. For this, we use the normalized B-splines $d_j^{-1/2}N_j$ which by virtue of (4.3) and (4.4) are an unconditional basis for $\mathcal{S}_r \cap L_2$. Therefore, if $f \in L_2$, then Pf is the unique spline in $\mathcal{S}_r$ which satisfies the orthogonality conditions $f - Pf \perp d_i^{-1/2}N_i$, $i \in \mathbb{Z}$. To each $f \in L_2$ corresponds a *moment sequence*

$$(4.5)\ \boldsymbol{\phi} := \boldsymbol{\phi}(f) := (\phi_i(f))_{i\in\mathbb{Z}}\,, \quad \phi_i(f) := \int f d_i^{-1/2} N_i dx = \int Pf d_i^{-1/2} N_i dx\,.$$

(But $\boldsymbol{\phi}$ does not determine f uniquely, only up to a summand $g \in L_2$ with $Pg = 0$.)

The projection P can now be completely described by the matrix

$$(4.6) \qquad A := A_2 := (a_{i,j})\,, \quad a_{i,j} := \int d_i^{-1/2} N_i d_j^{-1/2} N_j dx\,, \quad i,j \in \mathbb{Z}\,.$$

Namely, one has $Pf = \sum c_j d_j^{-1/2} N_j$ where the coefficients $c := (c_j) \in l_2$ satisfy the system of equations $Ac = \boldsymbol{\phi}$. Indeed, in terms of c, these equations

$$(4.7) \qquad \begin{aligned} (Ac)_i &= \sum_{j\in\mathbb{Z}} a_{i,j} c_j = \sum_{j\in\mathbb{Z}} \int d_i^{-1/2} N_i d_j^{-1/2} N_j c_j dx \\ &= \int d_i^{-1/2} N_i Pf dx = \phi_i(f) \end{aligned}$$

are equivalent to the orthogonality conditions $f - Pf \perp d_i^{-1/2} N_i$. Since Pf is uniquely determined by the orthogonality conditions, the system of equation (4.5) has a unique solution c.

Theorem 4.1 that follows shows that the norm of P on L_2 is equivalent to the norm of A_2^{-1} on $l_2(\mathbb{Z})$. On the other hand, the mapping properties of P on the other L_p spaces, $p \neq 2$, are best analyzed by using a different normalization for the B-spline basis. Namely, for $f \in L_2$, the projection Pf can also be described by $Pf = \sum c_j^{(p)} d_j^{-1/p} N_j$ where $c_p := (c_j^{(p)})$ satisfies the system of equations:

$$(4.8) \qquad \begin{cases} A_p c_p = \boldsymbol{\phi}_p \\ A_p := \left(a_{i,j}^{(p)}\right)\,, \quad a_{i,j}^{(p)} := \int d_i^{-1/p'} N_i \quad d_j^{-1/p} N_j dx\,. \\ \boldsymbol{\phi}_p := \boldsymbol{\phi}_p(f) := (\phi_i^{(p)}(f))\,, \quad \phi_i^{(p)}(f) := \int f d_i^{-1/p'} N_i\, dx \end{cases}$$

We obtain A_p from A_2 by pre- and post-multiplication by diagonal matrices.

Let us observe that *the matrix A_p is l_p bounded and has norm $\|A_p\|_p \le 1$.* To see this, for each $\boldsymbol{\alpha} = (\alpha_i) \in l_p$ and $\boldsymbol{\beta} = (\beta_i) \in l_{p'}$, we let $S := \sum \alpha_j d_j^{-1/p} N_j$, $R := \sum \beta_j d_j^{-1/p'} N_j$. Then, from (4.4), (4.8), and Hölder's inequality, we have

$$(4.9) \qquad \left|\sum \beta_i (A_p \alpha)_i\right| = \left|\int SRdx\right| \le \|S\|_p \|R\|_{p'} \le \|\boldsymbol{\alpha}\|_p \|\boldsymbol{\beta}\|_{p'}\,.$$

This shows that $\|A_p \boldsymbol{\alpha}\|_p \le \|\boldsymbol{\alpha}\|_p$, as required.

Likewise, for each $f \in L_p(\mathbb{R})$, the moment sequence $\boldsymbol{\phi}_p(f)$ is defined and maps L_p into l_p with norm ≤ 1:

$$(4.10) \qquad \|\boldsymbol{\phi}(f)\|_p \le \|f\|_p\,.$$

Indeed, if $\boldsymbol{\beta} := (\beta_i)_{i\in\mathbb{Z}} \in l_{p'}$ and $R := \sum \beta_i d_i^{-1/p'} N_i$, then

$$(4.11) \qquad \left|\sum \beta_i \phi_i(f)\right| = \left|\int fR\, dx\right| \le \|f\|_p \|R\|_{p'} \le \|f\|_p \|\boldsymbol{\beta}\|_{p'}\,.$$

Let Y be the kernel of the L_2 spline projection P. That is, Y is the set of $f \in L_2$ with $Pf = 0$. Equivalently, Y is the orthogonal complement of $\mathcal{S}_r \cap L_2$ in L_2. A function $f \in L_2(\mathbb{R})$ is in Y if and only if $\phi_i(f) = 0$, $i \in \mathbb{Z}$. We wish to describe the linear functionals $\lambda \in Y^\perp$ which are defined on L_2 and are orthogonal to Y. We shall show that they are all given by the formula

$$(4.12) \qquad \lambda = \sum \beta_i \phi_i\,, \quad (\beta_i) \in l_2\,,$$

where the ϕ_i are the moment functionals of (4.5). To see this, we first note that any linear functional of the form (4.12) can be written as $\lambda(f) = \int fR\, dx$ with R as in (4.11). By (4.11), it follows that $\|\lambda\| = \|R\|_2 \sim \|\boldsymbol{\beta}\|_2$. Therefore, the functionals (4.12) are bounded on L_2 and form a closed linear subspace Λ of the dual space of L_2. Now, if a functional λ_0 defined on L_2 and orthogonal to Y were not in Λ, then by Theorem 1.4 of Chapter 3, we could separate λ_0 and Λ by a function $g \in L_2$ so that $\lambda_0(g) > 0$, $\phi_i(g) = 0$, $i \in \mathbb{Z}$. But the latter equations would imply that $g \in Y$ and we obtain the contradiction $\lambda_0(g) = 0$.

A linear operator U defined on a space L_q has a unique continuous extension to L_p, $p \neq q$, if and only if for some constant $C > 0$, we have $\|Uf\|_p \le C\|f\|_p$ for all $f \in L_p \cap L_q$. The smallest C is the norm $\|U\|_p$ of U as a mapping on L_p. There is a (unique) continuous inverse U^{-1} for U (defined on the range of U) if and only if for some constant $C > 0$, we have $\|Uf\|_p \ge C\|f\|_p$, $f \in L_p \cap L_q$. For the largest such $C > 0$, the number C^{-1} gives the norm $\|U^{-1}\|_p$. The same remarks hold for linear operators on the l_p spaces and therefore in particular to bi-infinite matrices.

Theorem 4.1 (de Boor [1976_2]). *For the orthogonal projection P of L_2 onto $\mathcal{S}_r \cap L_2$, we have*

$$(4.13) \qquad D_r^2 \|A_2^{-1}\|_2 \le \|P\|_2 = 1 \le \|A_2^{-1}\|_2\,.$$

If $1 \le p \le \infty$ and if the matrix A_p of (4.7) has an inverse A_p^{-1} which is l_p bounded, then P has a unique bounded extension to $L_p(\mathbb{R})$ and

$$(4.14) \qquad \|P\|_p \le \|A_p^{-1}\|_p\,.$$

Proof. (i) We assume that A_p^{-1} exists and is l_p bounded. Let $f \in L_p \cap L_2$. By (4.10), the moment sequence $\boldsymbol{\phi}_p := \boldsymbol{\phi}_p(f)$ is in l_p and $\|\boldsymbol{\phi}_p\|_p \le \|f\|_p$. Then $Pf = \sum \alpha_i d_i^{-\frac{1}{p}} N_i$ where $\boldsymbol{\alpha} = A_p^{-1}\boldsymbol{\phi}_p$. It follows that $\boldsymbol{\alpha} \in l_p$ and by (4.4) and (4.10),

$$\|Pf\|_p = \left\|\sum \alpha_j d_j^{-\frac{1}{p}} N_j\right\|_p \le \|\boldsymbol{\alpha}\|_p \le \left\|A_p^{-1}\right\|_p \|\boldsymbol{\phi}_p\|_p \le \left\|A_p^{-1}\right\|_p \|f\|_p\,.$$

From the remarks preceding this theorem, we have that P has a continuous extension to L_p and $\|P\|_p$ satisfies (4.14). This also establishes the right inequality in (4.13).

(ii) To prove the left inequality in (4.13), we assume that $\boldsymbol{\alpha} \in l_2(\mathbb{Z})$ and majorize $\|\boldsymbol{\alpha}\|_2$ by a multiple of $\|\boldsymbol{\eta}\|_2$ with $\boldsymbol{\eta} = A_2\boldsymbol{\alpha}$. The function $S := \sum \alpha_j d_j^{-\frac{1}{2}} N_j$ is in $L_2 \cap \mathcal{S}_r$ and $PS = P(S-f)$ for all $f \in Y$. Hence, by (4.4),

$$D_r\|\boldsymbol{\alpha}\|_2 \le \|S\|_2 = \|PS\|_2 \le \|P\|_2 \operatorname{dist}(S,Y)_{L_2} = \|P\|_2 \sup|\lambda(S)|$$

where the supremum is extended over all $\lambda \in Y^\perp$ of norm one. We represent λ as in (4.12), $\lambda = \sum \beta_i \phi_i$ with $\boldsymbol{\beta} := (\beta_i)$ in l_2. If $R := \sum \beta_i d_i^{-\frac{1}{2}} N_i$, then by (4.4), $R \in L_2$ and $\lambda(f) = \int fR\,dx$. Again by (4.4),

$$1 = \|\lambda\| = \|R\|_2 \ge D_r\|\boldsymbol{\beta}\|_2 .$$

Therefore,

$$|\lambda(S)| = \left|\sum \beta_i \eta_i\right| \le \|\boldsymbol{\beta}\|_2\|\boldsymbol{\eta}\|_2 \le D_r^{-1}\|\boldsymbol{\eta}\|_2 .$$

We have proved that

$$\|\boldsymbol{\alpha}\|_2 \le D_r^{-2}\|P\|_2\|A_2\boldsymbol{\alpha}\|_2 .$$

By the remarks preceding this theorem, A_2 is invertible on l_2 and $\|A_2^{-1}\|_2$ satisfies the left inequality in (4.13). □

Corollary 4.2. *The inverse A_2^{-1} exists and satisfies*

$$(4.15) \qquad 1 \le \left\|A_2^{-1}\right\|_2 \le D_r^{-2} .$$

Moreover, for the condition number of (1.9),

$$(4.16) \qquad \kappa_2(A_2) := \|A_2\|_2 \left\|A_2^{-1}\right\|_2 \le D_r^{-2} .$$

Theorem 4.1 shows that to find an upper bound for $\|P\|_p$, it is enough to find one for $\left\|A_p^{-1}\right\|_p$. The following theorem of Demko for banded matrices A is useful for this purpose. A finite or bi-infinite matrix $A := (a_{i,j})$ has *bandwidth* d if $a_{i,j} = 0$, $|i-j| \ge d$ and if d is the smallest integer with this property. We then say *A is banded.* This means that the nonzero entries $a_{i,j}$ are located on the main diagonal of A and on the $2d-2$ nearby diagonals parallel to it. The matrices A_p of (4.8) have bandwidth r.

Theorem 4.3 (Demko [1977]). *If a bi-infinite matrix A with bandwidth r has a bounded inverse $A^{-1} = (b_{i,j})$ on l_p and $\kappa = \kappa_p(A) := \|A\|_p\|A^{-1}\|_p$ is the condition number of A, then*

$$(4.17) \qquad \begin{aligned} &\text{(i)} && |b_{i,j}| \le c_0\lambda^{|i-j|} , \\ &\text{(ii)} && \left\|A^{-1}\right\|_q \le \frac{2c_0}{(1-\lambda)} , \quad 1 \le q \le \infty , \end{aligned}$$

with $c_0 := \lambda^{-2r}\|A^{-1}\|_p$ and $\lambda := \left(\frac{\kappa^s-1}{\kappa^s+1}\right)^{\frac{1}{2rs}}$ where $s := p$ for $1 \le p \le 2$ and $s := p'$ for $2 \le p \le \infty$.

Proof. We can assume that $1 \le p \le 2$. Indeed, if $p > 2$, then for the transposed matrix A^t, which satisfies $(A^t)^{-1} = (A^{-1})^t$, the hypotheses of the theorem are valid with p replaced by $p' < 2$ and with the same λ.

We shall estimate the entries $b_{i,0}$, $i \ge 0$, in column zero of $B := A^{-1}$. The estimates for $b_{i,0}$ when $i < 0$ and for other columns are the same. Let $\boldsymbol{\beta} := (b_{i,0})$ and for each $n = 1, 2, \dots$, let the sequence $\boldsymbol{\beta}_n := (\beta_n(i))$ be given by

$$\boldsymbol{\beta}_n(i) := \begin{cases} b_{i,0} , & i > 2rn \\ 0 , & i \le 2rn . \end{cases}$$

Since A is banded, it follows that $(A\boldsymbol{\beta}_n)_i = 0$, $i \le (2n-1)r$. Also, since $\boldsymbol{\beta}$ is a column of A^{-1}, we have $(A\boldsymbol{\beta}_n)_i = (A\boldsymbol{\beta})_i = 0$, $i \ge (2n+1)r$. In particular, $A\boldsymbol{\beta}_n$ and $A\boldsymbol{\beta}_{n+1}$ have disjoint supports, $n = 1, 2, \dots$. With all norms in l_p, we have

$$\begin{aligned} \|\boldsymbol{\beta}_n\|^p + \left\|\boldsymbol{\beta}_{n+1}\right\|^p &\le \|A^{-1}\|^p \left[\|A\boldsymbol{\beta}_n\|^p + \left\|A\boldsymbol{\beta}_{n+1}\right\|^p\right] \\ &= \|A^{-1}\|^p \left\|A\left(\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n+1}\right)\right\|^p \\ &\le \kappa^p \left\|\boldsymbol{\beta}_n - \boldsymbol{\beta}_{n+1}\right\|^p = \kappa^p \left(\|\boldsymbol{\beta}_n\|^p - \left\|\boldsymbol{\beta}_{n+1}\right\|^p\right) . \end{aligned}$$

This gives

$$(4.18) \qquad \left\|\boldsymbol{\beta}_{n+1}\right\| \le \left(\frac{\kappa^p - 1}{\kappa^p + 1}\right)^{\frac{1}{p}} \|\boldsymbol{\beta}_n\| \le \lambda^{2r}\|\boldsymbol{\beta}_n\| .$$

We also have $\|\boldsymbol{\beta}_1\| \le \|\boldsymbol{\beta}\| = \|A^{-1}e_0\| \le \|A^{-1}\|$ where $e_0 := (\delta_{i,0})$ (with the Kronecker's δ). Hence, iterating (4.18), we obtain

$$(4.19) \quad \left\|\boldsymbol{\beta}_{n+1}\right\| \le \lambda^{2r}\|\boldsymbol{\beta}_n\| \le \dots \le \lambda^{2rn}\|\boldsymbol{\beta}_1\| \le \lambda^{2rn}\left\|A^{-1}\right\| , \quad n = 0, 1, \dots .$$

Therefore, if $i \ge 2r$, then (i) follows from (4.19). On the other hand, if $i < 2r$, then $|b_{i,0}| \le \|\boldsymbol{\beta}\| \le \|A^{-1}\|$ and (i) follows in this case as well.

For (ii), we have $\sum_{j\in\mathbb{Z}} |b_{i,j}| \le c_0(1 + 2\lambda + 2\lambda^2 + \dots) \le \frac{2c_0}{(1-\lambda)}$. Hence, $\|A^{-1}\|_\infty \le \max_i \sum_{j\in\mathbb{Z}} |b_{i,j}| \le \frac{2c_0}{(1-\lambda)}$. Similarly, $\|A^{-1}\|_1 \le \frac{2c_0}{(1-\lambda)}$. By the interpolation Theorem 4.4 of Chapter 2, we have (ii). □

We can apply Theorem 4.3 to obtain concrete estimates for $\left\|A_p^{-1}\right\|_p$ and $\|P_T\|_p$ under the assumption (4.2), $d_i/d_j \le M$. First of all, from Corollary 4.2 and Theorem 4.3, we obtain $\|A_2^{-1}\|_p \le C_r$, $1 \le p \le \infty$, where C_r is a constant depending only on r. Next, we compare the coefficients of A_p with those of A_2:

$$a_{i,j}^{(p)} = d_i^{\frac{1}{2}-\frac{1}{p'}} a_{i,j}^{(2)} d_j^{\frac{1}{2}-\frac{1}{p}} .$$

For the inverse matrices $A_p^{-1} = \left(b_{i,j}^{(p)}\right)$, $A_2^{-1} = \left(b_{i,j}^{(2)}\right)$ this yields

$$b_{i,j}^{(p)} = d_i^{\frac{1}{p}-\frac{1}{2}} b_{i,j}^{(2)} d_j^{\frac{1}{p'}-\frac{1}{2}} = \left(\frac{d_i}{d_j}\right)^{\frac{1}{p}-\frac{1}{2}} b_{i,j}^{(2)} . \tag{4.20}$$

This means that we obtain A_p^{-1} from A_2^{-1} by pre- and post-multiplication by diagonal matrices. Since $(\frac{d_i}{d_j})^{\frac{1}{p}-\frac{1}{2}} \le M^{|\frac{1}{p}-\frac{1}{2}|}$, we obtain $\|A_p^{-1}\|_p \le M^{|\frac{1}{p}-\frac{1}{2}|}\|A_2^{-1}\|_p \le C_r M^{|\frac{1}{p}-\frac{1}{2}|}$. From Theorem 4.1, we obtain

Theorem 4.4. *If the knot sequence T satisfies (4.2), then the projection P_T of order r has a continuous extension to L_p and*

$$\|P_T\|_p \le C_r M^{|\frac{1}{p}-\frac{1}{2}|} . \tag{4.21}$$

In particular,

$$\|P_T\|_\infty \le C_r M^{\frac{1}{2}} . \tag{4.22}$$

The history of the last theorem is quite interesting. With quite different proofs, it was first established by Douglas, Dupont, and Wahlbin [1975] for knots with a bounded mesh ratio $\frac{t_{i+1}-t_i}{t_{j+1}-t_j} \le M$, then by de Boor [1976$_2$] under the assumption (4.2). As we have seen it follows easily from Theorem 4.1 by Demko's result.

§ 5. Interpolation on $\mathbb{R}$

Let T be a bi-infinite knot sequence and let $X := (x_i)$ be a bi-infinite sequence of interpolation points. For splines $S \in \mathcal{S}_r(T, \mathbb{R})$, we would like to solve the Lagrange interpolation problem

$$S(x_i) = y_i , \quad i \in \mathbb{Z} , \tag{5.1}$$

for a given sequence $Y := (y_i)$ of real (or complex) data.

The cases easiest to treat are when r is odd and the knots interlace with the interpolation points: $\cdots < x_{-1} < t_{-1} < x_0 < t_0 < x_1 < \ldots$ and when r is even and $x_i = t_i$, $i \in \mathbb{Z}$. For $r = 1$ and $r = 2$, this problem always has a unique solution. It is a step function in the first case; a broken line connecting the points (x_i, y_i) in the second.

For $r > 2$, the flavor of the problem (5.1) is quite different (and different from the finite interpolation problems of §9, Chapter 5). Solutions to (5.1) exist but there is no uniqueness.

One can easily describe all solutions to (5.1). If r is even, we let S on $[t_0, t_1]$ be an arbitrary polynomial P_0 of degree $< r$ for which $P_0(t_i) = y_i$, $i = 0, 1$. Then on $[t_1, t_2]$, S must be a polynomial $P_1 \in \mathbb{P}_{r-1}$ which is uniquely described by the values $P_1^{(i)}(t_1) = P_0^{(i)}(t_1)$, $0 \le i \le r-2$ and $P_1(t_2) = y_2$. Continuing in this way, we obtain all solutions of (5.1). If r is odd, the situation is the same except that P_0 on $[t_0, t_1]$ is restricted only by the one condition $P_0(x_1) = y_1$.

The splines $S_0 \in \mathcal{S}_r(T, \mathbb{R})$, which interpolate the zero data, $S_0(x_i) = 0$, $i \in \mathbb{Z}$, are called *null splines*. It follows from the foregoing for both r even and odd, that they form a subspace $\mathcal{N} := \mathcal{N}_r$ of $\mathcal{S}_r$ of dimension 2μ where $\mu := [(r-1)/2]$. The importance of the null splines S_0 is that if S is a solution to (5.1) then $S + S_0$ is also a solution.

If $X = \mathbb{Z}$, we have the cardinal spline interpolation discussed in the next section. With the help of §4, we shall treat here the case of odd degree splines $r = 2m$, $m = 1, 2, \ldots$, when $X = T$, that is, $x_i = t_i$, $i \in \mathbb{Z}$, and T has bounded *global mesh ratio:*

$$\sup_{i,j \in \mathbb{Z}} \frac{t_{i+1}-t_i}{t_{j+1}-t_j} \le M . \tag{5.2}$$

Our main question will be whether there is a solution S to (5.1) which reflects properties of Y. For example, if $Y \in l_p$, does there exist a solution S in $\mathcal{S}_r(T, \mathbb{R}) \cap L_p(\mathbb{R})$? If yes, is it unique? The latter would be the case if and only if $\mathcal{N} \cap L_p$ consists of just the zero spline.

These questions have a positive answer if and only if this is true with the knot sequence T replaced by $\alpha T := (\alpha t_i)$, $\alpha > 0$. Therefore, instead of (5.2), we assume that

$$1 \le \Delta t_i \le M , \quad i \in \mathbb{Z} . \tag{5.3}$$

We shall need the remarks following Theorem 6.1 of Chapter 5. If there is a function f with $f^{(m)} \in L_2(\mathbb{R})$ and $f(x_j) = y_j$, then there is a unique spline $S \in \mathcal{S}_{2m}(T, \mathbb{R})$ with $S(x_i) = y_i$, $i \in \mathbb{Z}$, and $S^{(m)} \in L_2(\mathbb{R})$. Moreover, $\|S^{(m)}\|_2 \le \|f^{(m)}\|_2$ for each such f.

We apply this to the sequence $Y_i := (\delta_{i,j})_{j=-\infty}^{+\infty}$, $i \in \mathbb{Z}$, where $\delta_{i,j}$ are the Kronecker deltas. Clearly, for some $C > 0$ depending only on r, there are functions f_i with $f_i^{(m)} \in L_2(\mathbb{R})$ and $f_i(x_j) = \delta_{i,j}$, $\|f_i^{(m)}\|_2 \le C$. Hence there are unique splines L_i – called *fundamental splines* – which satisfy $L_i(x_j) = \delta_{i,j}$, $i, j \in \mathbb{Z}$, with the minimal $\|L_i^{(m)}\|_2 \le C$. One of the important properties of the L_i is their exponential decay.

Theorem 5.1. *If $r = 2m$ is even and if T satisfies (5.3), then there are $C_0 > 0$ and $0 < \lambda < 1$ depending only on r and M such that for each $i \in \mathbb{Z}$,*

$$|L_i(x)| \le C_0 \lambda^{|j-i|} , \quad x \in [t_j, t_{j+1}] . \tag{5.4}$$

Proof. From (5.3), we have $1 \le d_j := \frac{t_{j+m}-t_j}{m} \le M$. We fix a value of $i \in \mathbb{Z}$. Then the spline $L_i^{(m)}$ belongs to $\mathcal{S}_m(T, \mathbb{R}) \cap L_2(\mathbb{R})$ and has therefore the representation

$$L_i^{(m)} = \sum \alpha_j d_j^{-\frac{1}{2}} N_j , \quad N_j(x) = N(x; t_j, \ldots, t_{j+m}) , \quad \boldsymbol{\alpha} = (\alpha_i) \in l_2 .$$

According to (2.2) of Chapter 5, for its moments we have

$$\phi_j := d_j^{-\frac{1}{2}} \int L_i^{(m)} N_j dx = m! d_j^{-\frac{1}{2}} [t_j, \dots, t_{j+m}] L_i, \quad j \in \mathbb{Z}. \tag{5.5}$$

From this and (7.7) of Chapter 4, we see that $\phi_j = 0$ if $j \notin [i-m, i]$. Moreover,

$$|\phi_j| \le d_j^{-\frac{1}{2}} \left\| L_i^{(m)} \right\|_2 \|N_j\|_2 \le C, \quad j \in \mathbb{Z}. \tag{5.6}$$

For the matrix A_2 of (4.6),

$$A_2 \boldsymbol{\alpha} = \boldsymbol{\phi}, \quad \boldsymbol{\alpha} = A_2^{-1} \boldsymbol{\phi}.$$

Therefore, $\boldsymbol{\alpha}$ is a linear combination of the columns of $A_2^{-1} = (b_{k,j})$,

$$\alpha_k = \sum_{j=-\infty}^{\infty} b_{k,j} \phi_j = \sum_{j=i-m}^{i} b_{k,j} \phi_j.$$

From (4.17) applied to A_2, we have $|b_{k,j}| \le c_0 \lambda^{|k-j|}$, $j, k \in \mathbb{Z}$. It follows therefore that the α_k also decrease exponentially:

$$|\alpha_k| \le C \lambda^{|k-i|}, \quad k \in \mathbb{Z}, \tag{5.7}$$

with C (here and later) depending only on r and M.

If $x \in [t_j, t_{j+1}]$, then only the terms with indices $k \in (j-m, j]$ can be different from zero in the sum $L_i^{(m)}(x) = \sum \alpha_k d_k^{-\frac{1}{2}} N_k(x)$. Since the N_k are nonnegative and sum to 1,

$$\sup_{t_j \le x \le t_{j+1}} \left| L_i^{(m)}(x) \right| \le \max_{j-m<k\le j} \left| \alpha_k d_k^{-\frac{1}{2}} \right| \le C \lambda^{|i-j|}.$$

We can use this to estimate L_i. Given $x \in [t_j, t_{j+1}]$, we choose k closest to j so that $i \notin [k, k+m]$. Let J be the smallest interval which contains $x, t_{k+1}, \dots, t_{k+m}$. The zero polynomial interpolates L_i at $t_{k+1}, \dots, t_{k+m}$. Since $|k-j| \le m$, the Newton remainder formula for interpolation ((7.3) of Chapter 4) gives

$$\begin{aligned} |L_i(x)| &\le |(x - t_{k+1}) \dots (x - t_{k+m})| \; |[t_{k+1}, \dots, t_{k+m}, x] L_i| \\ &\le C \sup_{u \in J} \left| L_i^{(m)}(u) \right| \le C_0 \lambda^{|j-i|}, \quad x \in [t_j, t_{j+1}]. \end{aligned} \tag{5.8}$$ □

Theorem 5.2 (de Boor [1976₃]. *If T satisfies* (5.2), *and $r = 2m$ is even, then for each $Y = (y_i) \in l_p$, $1 \le p \le \infty$, the interpolation problem* (5.1) *has a unique solution $S \in \mathcal{S}_r(T, \mathbb{R}) \cap L_p(\mathbb{R})$.*

Proof. We can assume that T satisfies (5.3). One solution to (5.1) is $S = \sum y_i L_i$. Since $Y \in l_p$ implies $Y \in l_\infty$, this series converges uniformly and

defines a spline $S \in \mathcal{S}_r(T, \mathbb{R})$. Let $\lambda_1 := \lambda^{\frac{1}{2}}$. For $x \in [t_j, t_{j+1}]$, by (5.4) and Hölder's inequality,

$$|S(x)| \le \sum_{-\infty}^{\infty} |y_i L_i(x)| \le C \sum_{i=-\infty}^{\infty} |y_i| \lambda^{|i-j|} \le C \left(\sum_{i=-\infty}^{\infty} |y_i|^p \lambda_1^{p|i-j|} \right)^{1/p}.$$

Hence from (5.3), we have

$$\|S\|_p^p = \sum \|S\|_p^p [t_j, t_{j+1}] \le C \sum_{j=-\infty}^{\infty} \sum_{i=-\infty}^{\infty} |y_i|^p \lambda_1^{p|i-j|} \le C \sum_{i=-\infty}^{\infty} |y_i|^p,$$

which shows that (5.1) has a solution in $\mathcal{S}_r(T, \mathbb{R}) \cap L_p$.

To ensure that S is the unique solution, we suppose that there is a second solution $S_1 \in \mathcal{S}_r(T, \mathbb{R}) \cap L_p$. Then $S_0 := S - S_1$ is a null spline, $S_0(t_j) = 0$, $j \in \mathbb{Z}$. By means of the divided difference representation, as in (5.5), we see that, if N_j are the B-splines of order m for T,

$$\phi_j := \int S_0^{(m)} d_j^{-\frac{1}{p'}} N_j dx = 0, \quad j \in \mathbb{Z}.$$

Since $S_0 \in L_p(\mathbb{R})$, (5.3) and the Markov type inequality ((1.9) of Chapter 5) imply that $S_0^{(m)}$ is also in $L_p(\mathbb{R})$ and therefore has the representation $S_0^{(m)} = \sum \alpha_j d_j^{-\frac{1}{p}} N_j$, $\boldsymbol{\alpha} = (\alpha_j) \in l_p$. For the matrix A_p of §4, $\boldsymbol{\alpha} = A_p^{-1} \boldsymbol{\phi} = 0$. Thus, $S^{(m)} \equiv 0$. Representing the spline S_0 by means of Newton's remainder formula for Lagrange interpolation, we prove as in (5.8) that $S_0 \equiv 0$. □

§6. Cardinal Spline Interpolation

This is spline interpolation at the integers as introduced by Schoenberg and developed in his monograph [A-1973]. The spline space is $\mathcal{S}_r := \mathcal{S}_r(T, \mathbb{R})$ with the knots $T := (t_j)_{-\infty}^{\infty}$ selected as follows. If the degree $r-1$, $r = 1, 2, \dots$, of the splines is odd, then $t_j := j$ and therefore the knots and the interpolation points coincide. For even degree, the knots are placed halfway between the interpolation points, $t_j := j + \frac{1}{2}$, $j \in \mathbb{Z}$. Schoenberg [A-1973] has shown that many of the problems of general bi-infinite interpolation have simple and elegant solutions in the cardinal case.

We shall use the following properties of the B-splines N:

$$\begin{cases} N(x; t_1, \dots, t_n) = N(x + a; t_1 + a, \dots, t_n + a) \\ N(x; t_1, \dots, t_n) = N(x'; t_1, \dots, t_n). \end{cases} \tag{6.1}$$

The second relation is true when the knots are equally distributed, $\Delta t_i =$ const, and x, x' are symmetric about the midpoint of the interval $[t_1, t_n]$.

We can treat even and odd degree simultaneously by using the centralized B-splines of order r:

$$\Lambda(t) := \Lambda_r(t) := N\left(t; -\frac{r}{2}, -\frac{r}{2}+1, \dots, \frac{r}{2}\right), \quad t \in \mathbb{R}. \tag{6.2}$$

Then, Λ is an *even function* on $\mathbb{R}$ and by Theorem 3.3 of Chapter 5 its integer translates are a basis for $\mathcal{S}_r$; each $S \in \mathcal{S}_r$ can be represented by a *cardinal series:*

$$S(x) = \sum c_j \Lambda(x-j), \quad x \in \mathbb{R}, \tag{6.3}$$

with sums here and later over the range $j \in \mathbb{Z}$. We say that S is a *cardinal spline interpolant* to $Y := (y_j)_{-\infty}^{\infty}$ if

$$S(i) = \sum c_j \Lambda(i-j) = y_i, \quad i \in \mathbb{Z}. \tag{6.4}$$

Some properties of cardinal spline interpolation follow from the previous section. For example, if $r > 2$, the interpolation problem (6.4) has infinitely many solutions; the dimension of the null space $\mathcal{N}_r$ is 2μ where $\mu := [(r-1)/2]$.

The collocation matrix

$$C_r^* := (\Lambda(i-j))_{i,j=-\infty}^{\infty} \tag{6.5}$$

is a special case of a Toeplitz matrix $A = (a_{i,j})$ which are characterized by the property $a_{i+1,j+1} = a_{i,j}$. Toeplitz matrices play an important role in analysis. The function $\rho(z) := \sum_j a_{0,j} z^j$ of the complex variable z is called the *symbol* of A. The symbol of C_r^* is

$$\rho(z) := \rho_r(z) := \sum \Lambda_r(-j) z^j = \sum_{|j|<r/2} \Lambda_r(-j) z^j. \tag{6.6}$$

The function ρ is self-reciprocating $\rho\left(\frac{1}{z}\right) = \rho(z)$, $z \neq 0$. Obviously, $\rho(z) > 0$, for all real $z \geq 0$.

We can describe the null space $\mathcal{N}_r$ in terms of the zeros γ of ρ. If γ is such a zero then $S(x) := \sum \gamma^j \Lambda(x-j)$ is obviously in $\mathcal{N}_r$. We shall see that null splines of this type are a basis for $\mathcal{N}_r$. This is proved by showing that ρ_r has 2μ distinct zeros in Lemma 6.1 below.

Closely related to $\rho_r(z)$ are the functions

$$\Phi_n(x; z) := \sum_j N_{j,n+1}(x) z^j = \sum_j N(x; j, \dots, j+n+1) z^j. \tag{6.7}$$

Using (6.1), we derive the functional equation of Φ_n:

$$\Phi_n(x+s; z) = z^s \Phi_n(x; z), \quad s \in \mathbb{Z}. \tag{6.8}$$

Another useful formula is the recurrence relation of the Φ_n,

$$\Phi_n(x; z) = \frac{1-z}{n} \frac{d}{dz} \Phi_{n-1}(x; z) + \left(\frac{x}{n} + \frac{n-x}{nz}\right) \Phi_{n-1}(x; z). \tag{6.9}$$

This follows from the recurrence relation (2.8) of Chapter 5 for the $N_{j,n+1}$:

$$\begin{aligned} N(x; j, \dots, j+n+1) &= \frac{x-j}{n} N(x; j, \dots, j+n) \\ &+ \frac{n+j+1-x}{n} N(x; j+1, \dots, j+n+1). \end{aligned}$$

Accordingly, $\Phi_n(x; z)$ can be expressed as a sum of two sums: in the second, we replace $j+1$ by j and obtain

$$\Phi_n(x; z) = \sum_j N_{j,n}(x) \left(\frac{1-z}{n} j z^{j-1} + \frac{xz+n-x}{nz} z^j\right)$$

which is (6.9).

The relation between ρ_r and Φ_n is as follows:

$$\begin{aligned} \rho_{r+1}(z) &= \sum_j N\left(-j; -\frac{r+1}{2}, \dots, \frac{r+1}{2}\right) z^j \\ &= \sum_j N\left(\frac{r+1}{2}; j, \dots, j+r+1\right) z^j = \Phi_r\left(\frac{r+1}{2}; z\right). \end{aligned} \tag{6.10}$$

Let $s := [(r+1)/2]$. Then, $s + (r+1)/2 = r + \theta$ where

$$\theta := \begin{cases} \frac{1}{2}, & \text{if } r \text{ is even} \\ 1, & \text{if } r \text{ is odd.} \end{cases} \tag{6.11}$$

Then $\rho_{r+1}(z) = z^{-s} P_r(z)$, where $P_r(z) := \Phi_r(r+\theta; z)$ is the polynomial of degree r,

$$P_r(z) := \sum_{j=0}^{r} N(r+\theta; j, \dots, j+r+1) z^j = z^s \rho_{r+1}(z). \tag{6.12}$$

All the coefficients of P_r are > 0 except for the constant term which is zero if $\theta = 1$, $r > 0$. Thus, P_r has a zero at $z = 0$ if r is odd.

The number θ of (6.11) depends on r. To avoid this, for each fixed $\theta \in (0, 1]$, we define the sequence of polynomials by the same formula

$$\Pi_{r,\theta}(z) := \Phi_r(r+\theta; z) = \sum_{j=0}^{r} N_{j,r+1}(r+\theta) z^j, \quad r = 0, 1, \dots. \tag{6.13}$$

For example, $\Pi_{0,\theta} = 1$, $0 < \theta < 1$, $\Pi_{0,1}$ is not defined, and $\Pi_{1,1} = z$.

Lemma 6.1. *For $0 < \theta < 1$, the polynomials $\Pi_{r,\theta}$ have r negative zeros, for $\theta = 1$, they have $r-1$ negative zeros and the zero at $z = 0$.*

Proof. First we derive a recurrence relation for the $\Pi_r := \Pi_{r,\theta}$. It can be obtained from (6.9). For $n = r$, $x = r+\theta$, $\Phi_n(x; z) = \Pi_r(z)$, $\Phi_{n-1}(x; z) = \Phi_{r-1}(r+\theta; z) = z\Phi_{r-1}(r-1+\theta; z) = z\Pi_{r-1}(z)$, and therefore from (6.9)

$$\Pi_r(z) = \frac{z(1-z)}{r} \Pi_{r-1}'(z) + \left(z + (1-\theta)\frac{1-z}{r}\right) \Pi_{r-1}(z). \tag{6.14}$$

For $\theta = 1$, we define $\Pi^*_{r-1}(z) := \Pi_r(z)/z$, $r \geq 1$. Then (6.14) yields

$$\Pi^*_{r-1}(z) = \frac{z(1-z)}{r}\Pi^{*\prime}_{r-2}(z) + \left(z + \frac{1-z}{r}\right)\Pi^*_{r-2}(z)\,. \tag{6.15}$$

Turning first to (6.14); for $0 < \theta < 1$, we note that $\Pi_r(0) > 0$ for all r; moreover that $\Pi'_{r-1}(\gamma)$ and $\Pi_r(\gamma)$ are of the opposite sign at a simple negative zero γ of Π_{r-1}. We make the induction assumption that $\gamma_{r-1} < \gamma_{r-2} < \cdots < \gamma_1 < 0$ are the simple zeros of Π_{r-1}. Then the sign of $\Pi'_{r-1}(\gamma_i)$ is $(-1)^{i-1}$, and that of $\Pi_r(\gamma_i)$ is $(-1)^i$. This gives $r-1$ zeros γ^*_i of Π_r which interlace the $\gamma_i : \gamma_{r-1} < \gamma^*_{r-1} < \cdots < \gamma_1 \leq \gamma^*_1 < 0$. There is a further zero γ^*_r because $\Pi_r(z)$ has sign $(-1)^r$ for $z \to -\infty$. For the Π^*_r, the argument is the same. □

Translating this in terms of P_{r-1} and $\mu = [(r-1)/2]$, we obtain

Theorem 6.2. *The function ρ_r has exactly 2μ simple negative zeros, and if γ is a zero so is γ^{-1}.*

For example, $\gamma = -1$ is not a zero of ρ_r.

Theorem 6.3. *A basis for the null space $\mathcal{N}_r$ of $\mathcal{S}_r(T, \mathbb{R})$ is given by the splines*

$$S_\gamma(x) = \sum_j \gamma^j \Lambda_r(x-j) \tag{6.16}$$

where the γ are the 2μ zeros of ρ_r.

Proof. Since S_γ vanishes at each integer, it is in $\mathcal{N}_r$. There are $2\mu = \dim \mathcal{N}_r$ distinct γ. If $S(x) = \sum_\gamma c_\gamma S_\gamma(x) = \sum_{j\in\mathbb{Z}} \left(\sum c_\gamma \gamma^j\right) \Lambda(x-j) = 0$, then $\sum_\gamma c_\gamma \gamma^j = 0$, $j \in \mathbb{Z}$, by Theorem 3.2 of Chapter 5 and therefore $c_\gamma = 0$ for all γ by Example 1 of §3, Chapter 3. Hence the S_γ are linearly independent and therefore are a basis for $\mathcal{N}_r$. □

In order to study the existence and uniqueness of the solution of the interpolation problem (6.4) in some space of sequences, we shall establish the existence of the inverse C_r^{*-1} of the matrix C^*_r and estimate its norm. This can be done because the matrix C_r^{*-1} is associated with $\Omega_r(z) := 1/\rho_r(z)$.

Let $\Gamma_r := \{-\gamma_1, \ldots, -\gamma_\mu\}$ be the zeros of ρ_r in $(-1, 0)$, so that $0 < \gamma_i < 1$. Since the B-splines $\Lambda_r(\cdot - j)$ are a partition of unity ((3.14) of Chapter 5), we have $\rho_r(1) = \sum \Lambda(-j) = 1$. Therefore,

$$\rho_r(z) = \prod_{\gamma\in\Gamma_r} \frac{1+\gamma z}{1+\gamma}\frac{1+\gamma/z}{1+\gamma}\,.$$

The function $\Omega_r(z) := 1/\rho_r(z)$ has a Laurent series which converges on some open annulus A_0 containing the unit circle:

$$\Omega_r(z) = \sum_{-\infty}^{\infty} \omega_r(n) z^n\,, \quad z \in A_0\,. \tag{6.17}$$

Properties of the Laurent series of Ω_r follow from those of the functions $\frac{1}{1+\gamma z}$ and $\frac{1}{1+\gamma/z}$. For example, if $\lambda := \max_i \gamma_i$, then

$$|\omega_r(n)| \leq C\lambda^{|n|}\,, \quad n = 0, \pm 1, \ldots\,, \tag{6.18}$$

for a constant $C > 0$ not depending on n.

Since $\Omega_r(z) = \Omega_r(1/z)$, we have $\omega_r(-n) = \omega_r(n)$, $n = 1, 2, \ldots$. Also,

$$(-1)^n \omega_r(n) > 0\,, \quad \text{for all } n\,. \tag{6.19}$$

Indeed, the coefficient of z^n in the Laurent series of $\frac{1}{1+\gamma z}$ is 0 when $n < 0$ and has sign $(-1)^n$ if the $n \geq 0$. A similar statement holds for the coefficients of the Laurent series of $\frac{1}{1+\gamma/z}$ from which (6.19) follows.

We form the bi-infinite Toeplitz matrix

$$B_r := (\omega_r(i-j))_{i,j=-\infty}^{\infty}\,. \tag{6.20}$$

Then $B_r = C_r^{*-1}$, because

$$\sum_k \omega_r(i-k)\Lambda_r(k-j) = \delta_{i,j}\,, \quad i, j \in \mathbb{Z}\,. \tag{6.21}$$

This follows from the identity $\Omega_r(z)\rho_r(z)z^j = z^j$ which holds for $z \in A_0$. Indeed, we write this in the form (using the fact that Λ_r is even)

$$\left(\sum_l \omega_r(l)z^l\right)\left(\sum_k \Lambda(k-j)z^k\right) = z^j\,.$$

Comparing the coefficients of z^i on both sides, we get (6.21).

We can determine exactly the norm of $\left\|C_r^{*-1}\right\|_p$ on l_p.

Theorem 6.4. *For the matrix C_r^{*-1} and $1 \leq p \leq \infty$, we have*

$$\left\|C_r^{*-1}\right\|_p = \frac{1}{\rho_r(-1)} = \prod_{\gamma\in\Gamma_r}\left(\frac{1+\gamma}{1-\gamma}\right)^2\,. \tag{6.22}$$

Proof. If $\boldsymbol{\alpha} \in l_p$, $1 \leq p \leq \infty$, then

$$(B_r\boldsymbol{\alpha})_i = \sum_j \omega_r(i-j)\boldsymbol{\alpha}_j = \sum_j \omega_r(j)\alpha_{i-j}\,.$$

Hence $B_r\boldsymbol{\alpha}$ is a linear combination of vectors $(\alpha_{i-j})_{j=-\infty}^{\infty}$ which all belong to l_p and have norm $\|\boldsymbol{\alpha}\|_p$. It follows that

$$\left\|C_r^{*-1}\right\|_p = \|B_r\|_p \leq \sum_j |\omega_r(j)| = \Omega_r(-1) = \frac{1}{\rho_r(-1)}\,, \quad 1 \leq p \leq \infty\,. \tag{6.23}$$

There is equality in (6.23). To see this, we first consider $p = 2$ and use the isometry between l_2 and $L_2(\mathbb{T})$ which assigns to each $\boldsymbol{\alpha} \in l_2$ the function $f_{\boldsymbol{\alpha}}(z) := \frac{1}{\sqrt{2\pi}} \sum_k \alpha_k z^k$, $z \in \mathbb{T}$; $\|\boldsymbol{\alpha}\|_2 = \|f\|_2$. Then $C_r^* \boldsymbol{\alpha}$ corresponds to $\frac{1}{\sqrt{2\pi}} \sum_k (C_r^* \boldsymbol{\alpha})_k z^k = \frac{1}{\sqrt{2\pi}} \sum_k \sum_j \alpha_j \Lambda(k-j) z^k = \rho_r(z) f_{\boldsymbol{\alpha}}(z)$. Hence,

$$\|C_r^{*-1}\|_2 = \sup_{f \in L_2(\mathbb{T})} \frac{\|f\|_2}{\|\rho_r f\|_2}. \tag{6.24}$$

If we take $f := X_I / |I|^{\frac{1}{2}}$, where I is a small interval (in $\mathbb{T}$) containing -1, we see that the right side of (6.24) is $\geq 1/\rho_r(-1)$. Hence, equality holds in (6.23) when $p = 2$. By the convexity inequality in the Riesz-Thorin interpolation theorem (Theorem 4.3 of Chapter 2), we must have equality for all $1 \leq p \leq \infty$ in (6.23). □

Corollary 6.5 (Schoenberg [A-1973], $1 \leq p < \infty$, Subbotin [1967], $p = \infty$). *For $Y \in l_p$, $1 \leq p \leq \infty$, and $r \geq 1$, there is a unique spline $S \in \mathcal{S}_r(T, \mathbb{R}) \cap L_p(\mathbb{R})$ which satisfies* (6.4); *furthermore*

$$\|S\|_p \leq \frac{1}{\rho_r(-1)} \|Y\|_p. \tag{6.25}$$

Proof. If $\boldsymbol{\alpha} = C_r^{*-1} Y$, then $S(x) := \sum_{-\infty}^{\infty} \alpha_j \Lambda_r(x - j)$ satisfies (6.4). From Theorem 6.4 and (1.6), we have

$$\|S\|_p \leq \|\boldsymbol{\alpha}\|_p = \left\|C_r^{*-1} Y\right\|_p \leq \frac{1}{\rho_r(-1)} \|Y\|_p.$$

A non-trivial linear combination of the splines (6.16) cannot belong to L_p. Indeed, for each of them, $|\gamma| \neq 1$ and $|S_\gamma(x)|$ behaves like $|\gamma|^x$ or $|\gamma|^{-x}$, for $x \to \pm\infty$. Since $\mathcal{N}_r \cap L_p$ consists only of the zero spline, S is unique. □

We can describe other sequence spaces which admit unique cardinal interpolants. For this purpose, we use the fundamental splines L_i, $i \in \mathbb{Z}$, which satisfy $L_i(j) = \delta_{i,j}$, $j \in \mathbb{Z}$. Since $Y_i := (\delta_{i,j})$ is in l_∞, there is a *unique* fundamental spline L_i which is bounded. By (6.21), its cardinal spline series is

$$L_i(x) = \sum_k \omega_r(i-k) \Lambda_r(x-k). \tag{6.26}$$

It follows from (6.26) and (6.18) that $L_i(x)$ decays exponentially:

$$|L_i(x)| \leq C \lambda^{|x-i|} \tag{6.27}$$

where $0 < \lambda < 1$ and $C > 0$ are constants depending only on r.

Theorem 6.6 (Schoenberg [A-1973]). *If the bi-infinite sequence $Y := (y_n)$ satisfies $|y_n| = \mathcal{O}(n^k)$, then there is a unique spline $S \in \mathcal{S}_r(T, \mathbb{R})$ which satisfies* (6.4) *and satisfies $|S(x)| = \mathcal{O}(|x|^k)$, $|x| \to \infty$.*

Proof. The spline $S(x) := \sum_n y_n L_n(x)$ satisfies (6.4). From (6.27),

$$|S(x)| \leq C \sum_n |n|^k \lambda^{|n-x|} \leq C|x|^k.$$

As in the proof of Corollary 6.5, we see that no nontrivial null spline S_0 satisfies $|S_0(x)| = \mathcal{O}(|x|^k)$ and therefore S is unique. □

§ 7. Approximation from Shift Invariant Spaces

We say that a space $\mathcal{S}$, that consists of functions on $\mathbb{R}$, is *shift invariant* (for integers) if together with any function $S \in \mathcal{S}$, the space $\mathcal{S}$ also contains its *shifts* $S(\cdot + j)$, $j \in \mathbb{Z}$. Shift invariant spaces occur frequently in approximation. The Schoenberg space $\mathcal{S}_r(\mathbb{Z}, \mathbb{R})$ of cardinal splines has this property. Many other (univariate and multivariate) spaces of spline functions are also shift invariant. Moreover, if φ is a compactly supported function, then the collection $\mathcal{S}(\varphi)$ of all functions

$$S(x) := \sum_{k \in \mathbb{Z}} c_j \varphi(x - j), \tag{7.1}$$

where (c_j) is any sequence of real numbers, is a shift invariant space. For any $x \in \mathbb{R}$, only a finite number of terms in the sum (7.1) can be non-zero and therefore this series is absolutely and uniformly convergent on each compact subset of $\mathbb{R}$.

The approximation properties of the shift invariant space $\mathcal{S}(\varphi)$ can be described by the zeros of the Fourier transform of φ. This was first noticed by Schoenberg [1946] who described the polynomials in $\mathcal{S}(\varphi)$. Later Strang and Fix [1973] described the approximation properties of $\mathcal{S}(\varphi)$. Their results extend readily to multivariate functions and have important applications in numerical analysis. We shall consider here a simple case of their theory where the function φ is univariate, belongs to $L_1(\mathbb{R})$, and has compact support contained in $[-L, L]$ where $L > 0$ is an integer.

Our analysis relies heavily on Fourier transforms. For a function $f \in L_1(\mathbb{R})$, its Fourier transform is given by

$$\widehat{f}(y) := \int_{\mathbb{R}} f(x) e^{-iyx} dx \tag{7.2}$$

(We shall only need the Fourier transform for compactly supported functions f.) Here are some simple properties of Fourier transforms:

$$[f(a + \cdot)]\widehat{\ }(y) = e^{iay} \widehat{f}(y); \quad [f(-\cdot)]\widehat{\ }(y) = \widehat{f}(-y). \tag{7.3}$$

If f has compact support then $\widehat{f}$ is infinitely differentiable and

$$D^k \widehat{f}(y) = [(-i\cdot)^k f]\widehat{\ }(y). \tag{7.4}$$

Combining (7.3) and (7.4), we obtain a formula which describes the Fourier transform of the function $t^j f(x-t)$:

$$[(\cdot)^j f(x-\cdot)]\widehat{\ }(y) = (-i)^{-j} D_y^j \left[e^{-ixy}\widehat{f}(-y)\right] . \tag{7.5}$$

For the functions φ, we have the Poisson summation formula which relates the values of φ at the integers with $\widehat{\varphi}$ at the integer multiples of 2π. Let (with sums and integrals without limits over $(-\infty,\infty)$)

$$f(x) := \sum_j \varphi(x+j) . \tag{7.6}$$

The sum in (7.6) converges, because only a finite number of terms are non-zero on each compact set, and represents a function of period one which is in $L_1[0,1]$. We represent the function f by its Fourier series, $f(x) = \sum_j c_j e^{2\pi i j x}$. Here,

$$\begin{aligned} c_k &= \int_0^1 f(x) e^{-2\pi i k x}\,dx = \sum_j \int_0^1 \varphi(x+j) e^{-2\pi i k x}\,dx \\ &= \sum_j \int_j^{j+1} \varphi(x) e^{-2\pi i k x}\,dx = \int_{\mathbb{R}} \varphi(x) e^{-2\pi i k x}\,dx = \widehat{\varphi}(2k\pi), \quad k \in \mathbb{Z}. \end{aligned}$$

If the Fourier series of f converges to $f(0)$ at $x = 0$ (for example if φ is continuous and of bounded variation or in a Lipschitz class), it yields the Poisson formula

$$\sum_j \varphi(j) = f(0) = \sum_j \widehat{\varphi}(2j\pi) . \tag{7.7}$$

The Poisson summation formula can be used to describe (in terms of $\widehat{\varphi}$) the polynomials in $\mathcal{S}$. As an example, we show that $(\varphi(x-j))_{k\in\mathbb{Z}}$ is a partition of unity on $\mathbb{R}$ if and only if

$$\widehat{\varphi}(0) = 1 \text{ and } \widehat{\varphi}(2j\pi) = 0 , \quad j \in \mathbb{Z}, \quad j \neq 0 . \tag{7.8}$$

Indeed, $\varphi(\cdot - j)$, $j \in \mathbb{Z}$, is a partition of unity if and only if the function f of (7.6) is identically one and this is true if and only if its Fourier coefficients $c_j = \widehat{\varphi}(2j\pi)$ satisfy (7.8).

To go further and describe the polynomials of higher degree which are in $\mathcal{S}$, we shall assume that φ is continuous of compact support and of bounded variation. It follows that f is continuous and of bounded variation on $[0,1]$ because only a finite number of $\varphi(\cdot - j)$, $j \in \mathbb{Z}$, are not identically zero on $[0,1]$. We begin with the following result of Schoenberg [1946].

Theorem 7.1. *If r is a nonnegative integer and φ is a continuous function of bounded variation with compact support satisfying*

$$\widehat{\varphi}(0) \neq 0 \text{ and } D^j\widehat{\varphi}(2k\pi) = 0 , \text{ for all } k \in \mathbb{Z},\ k \neq 0 \text{ and } j = 0,\dots,r , \tag{7.9}$$

then there are polynomials P_j of degree j such that

$$\sum_{k\in\mathbb{Z}} P_j(k)\varphi(x-k) = x^j , \quad j = 0,\dots,r . \tag{7.10}$$

Proof. By multiplying by a constant, we can assume that $\widehat{\varphi}(0) = 1$. Using the Leibniz differentiation formula in (7.5), we find

$$[(\cdot)^j\varphi(x-\cdot)]\widehat{\ }(2k\pi) = 0 , \quad k \in \mathbb{Z}\backslash\{0\} , \quad j = 0,\dots,r . \tag{7.11}$$

On the other hand for $k = 0$,

$$[(\cdot)^j\varphi(x-\cdot)]\widehat{\ }(0) = \widehat{\varphi}(0)x^j + R_{j-1}(x) = x^j + R_{j-1}(x) \tag{7.12}$$

where R_{j-1} is a polynomial of degree $< j$. Hence, from the Poisson summation formula for the function $t^j\varphi(x-t)$ (which is continuous and of bounded variation), we obtain,

$$\begin{aligned} \sum_k k^j\varphi(x-k) &= \sum_k [(\cdot)^j\varphi(x-\cdot)]\widehat{\ }(2k\pi) \\ &= x^j + R_{j-1}(x) , \quad j = 0,\dots,r . \end{aligned} \tag{7.13}$$

We can now prove (7.10) by induction on j. For $j = 0$, because of (7.8), we can take $P_0 \equiv 1$. Assume now that $s \le r$ and P_j has been constructed for $j = 0,\dots,s-1$. We write $R_{s-1}(x) =: c_0 + \cdots + c_{s-1}x^{s-1}$ and define $P := c_0P_0 + \cdots + c_{s-1}P_{s-1}$ and $P_s(x) := x^s - P(x)$. Then,

$$\begin{aligned} \sum_k P_s(k)\varphi(x-k) &= x^s + R_{s-1}(x) - \sum_k P(k)\varphi(x-k) \\ &= x^s + R_{s-1}(x) - R_{s-1}(x) = x^s . \end{aligned}$$

□

There is no strict converse to Theorem 7.1. For example if φ takes the value 1 on $[-1,0]$, -1 on $[0,1]$ and is otherwise zero, then it is easy to see that constants are in the space $\mathcal{S}$ but $\widehat{\varphi}(0) = 0$. Therefore, if $\mathcal{S}$ contains $\mathcal{P}_r$, we cannot conclude that $\widehat{\varphi}(0) \neq 0$. However, de Boor [1987] has shown that all other conditions of (7.9) must be satisfied. To see this, we shall use the following simple remark. If $S \in \mathcal{S}$, $S = \sum c_\nu\varphi(x-\nu)$, then by a change of summation order, we obtain

$$\begin{aligned} \sum_j S(j)\varphi(x-j) &= \sum_j\sum_\nu c_\nu\varphi(j-\nu)\varphi(x-j) \\ &= \sum_l\sum_\nu c_\nu\varphi(l)\varphi(x-\nu-l) = \sum_l \varphi(l)S(x-l) . \end{aligned} \tag{7.14}$$

In particular, if P is a polynomial in $\mathcal{S}$ of degree k, then since φ has compact support, (7.14) shows that $\sum_j P(j)\varphi(x-j)$ is a polynomial of degree $\le k$.

Theorem 7.2. *If the space $\mathcal{S}(\varphi)$ contains all polynomials of degree $\le r$, then*

$$D^j\widehat{\varphi}(2k\pi) = 0 , \quad k \neq 0 ,\ j = 0,\dots,r . \tag{7.15}$$

Proof. We shall prove the theorem by induction on r. For $r = 0$, we have that all constants are in $\mathcal{S}$ and therefore by (7.14), the function $f(x) := \sum_\nu \varphi(x+\nu)$ of (7.6) is a constant. Therefore its Fourier coefficients $c_k = \widehat{\varphi}(2\pi k)$, are 0 if $k \neq 0$. Assume now that the theorem has been proven for $r-1$. For the function $\psi(t) := t^r\varphi(x-t)$, we have by (7.14) and the Poisson summation formula (7.7),

$$Q(x) = \sum_k \psi(k) = \sum_k \widehat{\psi}(2k\pi) \tag{7.16}$$

with Q a polynomial of degree $\leq r$. Now, we can compute $\widehat{\psi}(2k\pi)$ by using Leibniz's formula in (7.5). It gives that $\widehat{\psi}(0)$ is an algebraic polynomial in x of degee $\leq r$. For other values of k, we have $D^\nu\widehat{\varphi}(2k\pi) = 0$, $k \neq 0$, $\nu = 0, \dots, r-1$ because of our induction hypothesis. From this it follows that $\widehat{\psi}(2k\pi) = i^{-r}D^r\widehat{\varphi}(-2k\pi)e^{-2\pi ikx}$. Using this in (7.16) shows that $\sum_{k\neq 0} D^r\widehat{\varphi}(2\pi k)e^{-2\pi ikx}$ is an algebraic polynomial. Now, this series converges to a function of period 1 because its coefficients are a multiple of $\widehat{\psi}(2\pi k)$ and the Fourier series of ψ converges at 0. Therefore, we must have that the sum of this series is a constant and all coefficients in this series are zero. This verifies that $D^r\widehat{\varphi}(2k\pi) = 0$, $k \neq 0$. □

Here are some examples of functions φ which satisfy (7.9). From (2.10) of Chapter 5, the B-spline $N_r(x) := N(x; 0, \dots, r)$ is the r-fold convolution of the characteristic function χ of $[0,1]$ with itself. From (7.2), $\widehat{\chi}(y) = (1-e^{-iy})/(iy)$, and hence

$$\widehat{N}_r(y) = \left(\frac{1-e^{-iy}}{iy}\right)^r . \tag{7.17}$$

Therefore N_{r+1} satisfies (7.9). Of course we could also derive (7.9) from Theorem 7.2. We can obtain functions φ which satisfy (7.9) by convolution with the N_r. For example, if $g \in L_1(\mathbb{R})$ and $\widehat{g}(0) = 1$, then $\varphi = g * N_{r+1}$ has Fourier transform $\widehat{\varphi} = \widehat{g}\widehat{N}_{r+1}$ and therefore φ satisfies (7.9) provide $\widehat{g}$ is sufficiently differentiable.

We assume for the remainder of this section that φ satisfies (7.10). We shall construct quasi-interpolant operators for approximating functions $f \in L_p$ by the elements of the dilated space $\mathcal{S}_h := \{S(\cdot/h) : S \in \mathcal{S}\}$, $h > 0$. If $P \in \mathcal{P}_r$, then the sum $\sum_k P(k)\varphi(x-k)$ is a polynomial in $\mathcal{P}_r$. Since the polynomials P_j of Theorem 7.1 are a basis for $\mathcal{P}_r$, if $P, Q \in \mathcal{P}_r$ and

$$\sum_k P(k)\varphi(x-k) = \sum_k Q(k)\varphi(x-k), \quad x \in \mathbb{R}, \tag{7.18}$$

then $P = Q$. From this we shall derive that each polynomial $P \in \mathcal{P}_r$ has a unique representation:

$$P(x) = \sum_k a_k(P)\varphi(x-k) \tag{7.19}$$

with

$$a_k(P) := \sum_{j=0}^r \frac{P^{(j)}(k)}{j!}P_j(0) = a_0(P(\cdot + k)) . \tag{7.20}$$

Indeed, for $m \in \mathbb{Z}$, and $j = 0, \dots, r$, we have $(x-m)^j = \sum_k P_j(k)\varphi(x-m-k) = \sum_k P_j(k-m)\varphi(x-k)$. This and the Taylor expansion of P about m yield

$$P(x) = \sum_k \sum_{j=0}^r \frac{P^{(j)}(m)}{j!}P_j(k-m)\varphi(x-k) . \tag{7.21}$$

The interior sum $\sigma_m(x) := \sum_{j=0}^r \frac{P^{(j)}(m)}{j!}P_j(x-m)$ is a polynomial in $\mathcal{P}_r$. From (7.18), we derive that all σ_m are identical. Replacing $\sigma_m(k)$ by $\sigma_k(k)$ in (7.21), we obtain (7.20).

We use the representation (7.19) to construct quasi-interpolant operators Q_h. By the norm comparison Theorem 2.7 of Chapter 4, we have that for any interval Ω containing 0 in its interior,

$$|a_0(P)| \leq C\|P\|_1(\Omega), P \in \mathcal{P}_r ,$$

where the constant C depends only on Ω, φ and r. By the Hahn-Banach theorem, we extend a_0 from $\mathcal{P}_r$ to $L_1(\Omega)$. In this way, for some bounded function G on Ω, we obtain

$$a_0(f) = \int_\Omega f(t)G(t)dt, \quad f \in L_1(\Omega), \quad \|G\|_\infty \leq C. \tag{7.22}$$

We extend G to $\mathbb{R}$ by putting $G(t) = 0$, $t \notin \Omega$. Formula (7.22) induces a representation of the functionals a_k. For each f that is locally integrable on $\mathbb{R}$,

$$a_k(f) = a_0(f(\cdot + k)) = \int_\Omega f(t+k)G(t)dt = \int_{\mathbb{R}} f(t)G(t-k)dt. \tag{7.23}$$

In what follows, we shall take $\Omega := [-L, L]$ although any other choice would work equally well. If $\Omega_k := \Omega + k = [-L+k, L+k]$, $k \in \mathbb{Z}$, then

$$|a_k(f)| \leq C\|f\|_1(\Omega_k) .$$

We now define the quasi-interpolant operator Q_h by

$$Q_h(f, x) := \sum_{k\in\mathbb{Z}} a_k(f(h\cdot))\,\varphi\left(\frac{x}{h} - k\right) . \tag{7.24}$$

Clearly,

$$Q_h(P) = P, \quad P \in \mathcal{P}_r . \tag{7.25}$$

Substituting (7.23) into (7.24), we obtain for each locally integrable f,

$$(7.26)\qquad \begin{aligned} Q_h(f,x) &= \frac{1}{h}\int_{\mathbb{R}} f(t)\sum_k G\left(\frac{t}{h}-k\right)\varphi\left(\frac{x}{h}-k\right)dt \\ &= \int_{\mathbb{R}} f(t)K_h(t,x)dt\,, \end{aligned}$$

where

$$(7.27)\qquad K_h(t,x) := \frac{1}{h}\sum_k G\left(\frac{t}{h}-k\right)\varphi\left(\frac{x}{h}-k\right), \quad t,x\in\mathbb{R}\,.$$

Since both G and φ are supported on Ω, the k-th term in (7.27) is non-zero only if x/h and t/h are in Ω_k. This can occur for at most $2L$ values of k. Hence, with $M := \|\varphi\|_\infty$,

$$(7.28)\qquad |K_h(t,x)| \begin{cases} =0\,, & \text{if } |x-t|\ge 2Lh\,,\\ \le 2LCMh^{-1}\,, & x,t\in\mathbb{R}\,. \end{cases}$$

It follows that $\int_{\mathbb{R}}|K_h|dt \le 4L^2CM =: C_1$ with a similar estimate for $\int_{\mathbb{R}}|K_h|dx$. We obtain

Theorem 7.3. *The quasi-interpolant Q_h reproduces polynomials of degree $\le r$; it has the representation* (7.26) *with the kernel K_h of* (7.27) *satisfying*

$$(7.29)\qquad \int |K_h|dt \le C_1\,, \quad \int |K_h|dx \le C_1\,.$$

It follows that Q_h is a bounded mapping on L_1 and L_∞ with norm $\le C_1$.

From the interpolation Theorem 4.4 of Chapter 2, we derive that Q_h maps each rearrangement invariant function space $X(\mathbb{R})$ into itself with norm $\le C_1$. For example for the space $L_p(\mathbb{R})$, we obtain

Theorem 7.4. *Let φ be a continuous function of bounded variation and compact support which satisfies* (7.9) *for some integer $r-1$ (in place of r). If $1\le p\le\infty$ and $h>0$, then Q_h boundedly maps L_p into itself. If $f\in W_p^r$, then*

$$(7.30)\qquad \|f-Q_h(f)\|_p \le C\,h^r\|f^{(r)}\|_p\,;$$

and moreover, if $f\in L_p(\mathbb{R})$,

$$(7.31)\qquad \|f-Q_h(f)\|_p \le C\omega_r(f,h)_p$$

where C is a constant depending only on φ and r.

Proof. In the case of (7.30), we use the Taylor polynomial $P := P(x,\cdot)$ of degree $r-1$ for f at x and its remainder $R(x,t) := f(t)-P(x,t)$. Since $P(x)=f(x)$ and Q_h reproduces polynomials, we have for fixed $x\in\mathbb{R}$,

$$(7.32)\qquad |f(x)-Q_h(f,x)| = |Q_h(f-P,x)| = \left|\int K_h(t,x)R(x,t)dt\right|\,.$$

By (7.28), $K_h(x,t)=0$ if $|x-t|\ge 2Lh$, and by (5.6) of Chapter 2, for $|x-t|\le 2Lh$,

$$|R(x,t)| \le \frac{(4Lh)^{r-1}}{(r-1)!}\int_{x-2Lh}^{x+2Lh}\left|f^{(r)}(u)\right|du \le C\,h^{r-1/p}\left(\int_{x-2Lh}^{x+2Lh}\left|f^{(r)}\right|^p du\right)^{1/p}.$$

We replace R by this and K_h by C_1/h in (7.32), integrate with respect to t for $|x-t|\le 2Lh$, and then take an L_p norm with respect to x to obtain (7.30). Then (7.31) follows by an application of Theorem 5.1 of Chapter 7. □

§ 8. Shape Preserving Interpolation

In this and the next section, we shall describe Lagrange interpolation for points $0\le x_0 < x_1\cdots < x_n \le 1$:

$$(8.1)\qquad S(x_i) = y_i\ (= f(x_i))$$

by spline functions S which preserve the shape of the data $Y := (y_i)$ (or of the function f). For example, if the function f is monotone or convex on some interval $[x_j,x_k]$, we would like to have an S which also has these properties. To achieve this, we take splines S with knots at the points x_i, which have more parameters than necessary to satisfy (8.1). The additional parameters are then selected to ensure the desired properties of S.

In the present section, the splines are C^1 cubics and the additional parameters are the slopes of S at the x_i; in the next section, the splines are C^1 quadratics and the additional parameters are simple knots distinct from the x_i. The spline S can then be represented by one of the known formulas, for example by (3.3) for cubic splines. The main point, however, is to see whether the approximation error $\|f-S\|$ remains small under the proposed algorithms which describe S.

For cubic splines, it is first necessary to determine when a cubic polynomial P is monotone on an interval, which we for the moment select to be $[0,1]$. We do this in terms of the values $y_0 := P(0)$, $y_1 := P(1)$, $s_0 := P'(0)$ and $s_1 := P'(1)$. We let M be the region pictured in Figure 8.1. Its boundary consists of the line segments $[(0,0),(0,3)]$ and $[(0,0),(3,0)]$ and the indicated portion of the ellipse $\Phi(x,y)=0$ with $\Phi(x,y) := (x-1)^2+(x-1)(y-1)+(y-1)^2-3(x+y-2)$. This ellipse is tangent to the coordinate axis at $(3,0)$ and $(0,3)$. By aM we denote the set of (ax,ay) with $(x,y)\in M$.

The region M is the disjoint union of subregions M_0, M_1, M_2, M_3 described as the set of all (x,y) in the first quadrant satisfying:

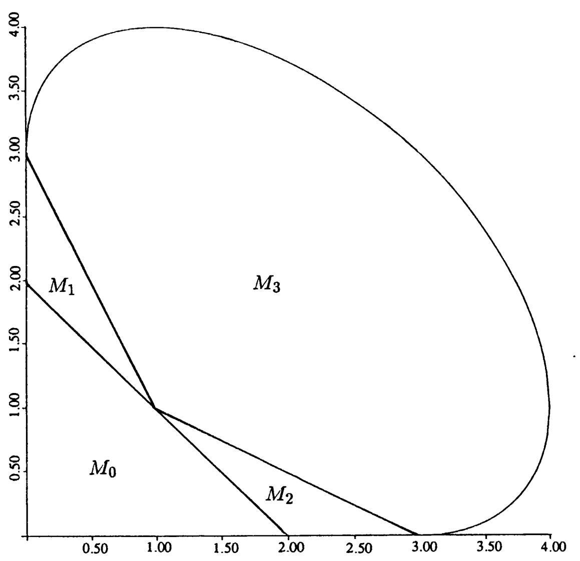


Fig. 8.1. Monotonicity Region for Cubics

$$
(8.2) \qquad
\begin{aligned}
&M_0 : x+y-2 \le 0 \\
&M_1 : x+y-2 > 0\,, \quad 2x+y-3 \le 0\,, \\
&M_2 : x+y-2 > 0\,, \quad x+2y-3 \le 0\,, \\
&M_3 : x+2y-3 > 0\,, \quad 2x+y-3 > 0\,, \quad \Phi(x,y) \le 0\,.
\end{aligned}
$$

Lemma 8.1 (Fritsch-Carlson [1980]). *A cubic polynomial P is monotone on $I = [0,1]$ if and only if $(s_0, s_1) \in \delta M$ for $\delta := y_1 - y_0$.*

Proof. If $\delta = 0$, then P is monotone if and only if it is constant which is equivalent to $s_0 = s_1 = 0$. We can therefore assume that $\delta \ne 0$. By considering $-P$ in place of P, if necessary, we can assume that $\delta > 0$. We must then have $s_0, s_1 \ge 0$, i.e. (s_0, s_1) must be in the first quadrant.

From the Newton representation ((6.2) of Chapter 4), one checks that P is given by

$$
(8.3) \qquad P(x) = y_0 + s_0 x + (\delta - s_0)\,x^2 + (s_0 + s_1 - 2\delta)\,x^2(x-1)\,.
$$

We consider the following two possibilities for $a := s_0 + s_1 - 2\delta$. If $a \le 0$, then the quadratic P' is concave and hence $P'(x) \ge \min(s_0, s_1) \ge 0$. Thus, in this case, $(s_0, s_1) \in \delta M_0$ is trivially necessary and sufficient for P to be monotone.

If $a > 0$, then P' has a unique extremum at $\xi := (2s_0 + s_1 - 3\delta)/3a$ and

$$
P'(\xi) = \delta \Psi(s_0/\delta,\ s_1/\delta)\,, \qquad \Psi(x,y) := -\Phi(x,y)/3(x+y-2)\,.
$$

Hence, P will be monotone on I if and only if either (a) $\xi \notin (0,1)$ or (b) $\xi \in (0,1)$ and $P'(\xi) \ge 0$. Case (a) occurs if and only if either $2s_0 + s_1 \le 3\delta$ or $s_0 + 2s_1 \le 3\delta$, i.e. if and only if $(s_0, s_1) \in \delta M_1$ or $(s_0, s_1) \in \delta M_2$. Case (b) holds if and only if $(s_0, s_1) \in \delta M_3$. $\square$

Suppose now that we are given points (x_i, y_i), $i = 0, \dots, n$, with $0 =: x_0 < x_1 < \cdots < x_n := 1$. We let $\mathcal{S} := \mathcal{S}_r(T, [0,1])$ where $T := (t_i)$, with $t_{2i-1} = t_{2i} = x_i$, $i = 1, \dots, n-1$, are the basic knots. Then $\dim(\mathcal{S}) = 2n+2$; the splines $S \in \mathcal{S}$ are C^1 cubics. A spline $S \in \mathcal{S}$ satisfying (8.1) can be represented by (3.3) where the y_i, $i = 0, \dots, n$, are the given values and the slopes s_i, $i = 0, \dots, n$, are free parameters. We would like to choose the slopes so that S is comonotone with (y_i). If $y_j, \dots, y_k$ is a nondecreasing (or nonincreasing) sequence then we want S to be likewise nondecreasing (nonincreasing) on $[x_j, x_k]$. The preceding lemma shows that this will be the case if and only if for $\delta_i := \Delta y_i / \Delta x_i$, we have

$$
(8.4) \qquad (s_i, s_{i+1}) \in \delta_i M\,, \quad i = 0, 1, \dots, n-1\,.
$$

Fritsch and Carlson [1980] have suggested the following general algorithm for defining S:

(8.5)

- (i) Assign slopes s_i, $i = 0, \dots, n$,
- (ii) Replace s_i by 0 if $s_i \delta_{i-1} \le 0$ or if $s_i \delta_i \le 0$,
- (iii) Modify the s_i, $i = 0, \dots, n$ so that $(s_i,\ s_{i+1}) \in \delta_i M$, $i = 0, \dots, n-1$.

The initial assignment of slopes is typically made by using a formula for numerical differentiation. For example, given points $a < b < c$, we can use the three-point formula

$$
(8.6) \quad D_x f := \alpha[a,b]f + \beta[b,c]f\,, \quad \beta := (2x - a - b)/(c-a)\,, \quad \alpha := 1 - \beta
$$

for approximating $f'(x)$ in terms of the values of f at a, b, c. Its justification lies in the fact that $D_x Q = Q'(x)$ for each quadratic polynomial Q. For any $x \in [a,c]$ and the supremum norm

$$
(8.7) \qquad |f'(x) - D_x f| \le 2(c-a)^2 \|f^{(3)}\|[a,c]\,.
$$

Indeed, if Q is the quadratic Taylor polynomial for f at x, then

$$
(8.8) \qquad f'(x) - D_x f = (f'(x) - Q'(x)) + (Q'(x) - D_x Q) + D_x(Q - f)\,.
$$

The first two differences on the right side of (8.8) are 0 while the third in absolute value for some $\xi, \eta \in [a,c]$ does not exceed

$$
\begin{aligned}
|\alpha(f-Q)'(\xi) + \beta(f-Q)'(\eta)| &\le (|\alpha| + |\beta|)\|f' - Q'\|[a,c] \\
&\le (|\alpha| + |\beta|)\|f^{(3)}\|[a,c](c-a)^2/2 \\
&\le 2\|f^{(3)}\|(c-a)^2
\end{aligned}
$$

because each of α and β is less than 2 in absolute value.

We shall use $D_x f$ to assign slopes s_i in step (i) as follows:

$$s_i := D_{x_i} f\,, \tag{8.9}$$

with $a_i := x_{i-1}$, $b_i := x_i$, $c_i = x_{i+1}$, if $0 < i < n$; $a_0 := x_0$, $b_0 := x_1$, $c_0 := x_2$, if $i = 0$; $a_n := x_{n-2}$, $b_n := x_{n-1}$, $c_n := x_n$, if $i = n$. Then, with $h_i := c_i - a_i$, we have from (8.7)

$$|f'(x_i) - s_i| \le 2h_i^2 \|f^{(3)}\|\,[a_i, c_i]\,, \quad i = 0, \dots, n\,. \tag{8.10}$$

The core of the algorithm (8.6) lies in how (iii) is implemented. Changing the value of s_i effects two points, (s_{i-1}, s_i), (s_i, s_{i+1}), and it may happen that a change in s_i which moves one of these points into its monotonicity region will move the other point out of its monotonicity region. One way of guaranteeing that this does not happen is to choose a subregion N of M with the property that $(x, y) \in N$ and $0 \le x' \le x$, $0 \le y' \le y$ imply $(x', y') \in N$. Then (iii) can be accomplished by decreasing the absolute values of the s_i until $(s_i, s_{i+1}) \in \delta_i N$, $i = 0, \dots, n-1$.

As an example, we consider the square $N := [0,3] \times [0,3]$ which is contained in the region M. To implement (iii), we follow the prescription

(8.11) For each i satisfying $|s_i| > 3\min(|\delta_{i-1}|, |\delta_i|)$, one replaces s_i by $3\min(|\delta_{i-1}|, |\delta_i|)\operatorname{sign} s_i$,

where $\delta_{-1} := \delta_n := \infty$ for the purpose of this formula. Then $(s_i, s_{i+1}) \in \delta_i N$, $i = 0, \dots, n-1$, and hence the interpolating spline S will be comonotone with the data (y_i). If the data are values of a function f, $y_i := f(x_i)$, then S is an approximant to f which is monotone on $[x_j, x_k]$ whenever f is.

The following theorem is an improvement on the error estimate of Eisenstadt, Jackson and Lewis [1985].

Theorem 8.2. *Let $f \in C^3(I)$, $I := [0,1]$, be monotone on each interval $[x_i, x_{i+1}]$, $i = 0, \dots, n-1$. Then the cubic spline interpolant S constructed by means of the algorithm* (8.5) *with* (i) *given by* (8.9) *and* (iii) *given by* (8.11) *satisfies*

$$\|f - S\| \le C\|f^{(3)}\| h^3\,, \quad h := \max \Delta x_i \tag{8.12}$$

where C is an absolute constant.

Proof. Let $\varepsilon_i := \|f^{(3)}\|\,[a_i, c_i]\,h_i^2$. We shall show that the final slope s_i produced by the algorithm satisfies

$$|s_i - f'(x_i)| \le 2\varepsilon_i\,. \tag{8.13}$$

For the value of s_i after step (i), this is true because of (8.10). This estimate is maintained at step (ii) even if s_i is changed because the new s_i approximates $f'(x_i)$ better than the old. For example, suppose $s_i > 0$ after step (i) is changed at step (ii). Then either $\delta_i \le 0$, or $\delta_{i-1} \le 0$. In both cases, $f'(x_i) \le 0$ and

therefore the new slope $s_i' = 0$ after step (iii) is a better approximation to $f'(x_i)$ than s_i.

We now suppose that the s_i at the beginning of step (iii) have been selected in accordance with (8.9) and (ii) and that (8.11) replaces s_i by s_i'. We shall assume that $\delta_i \ge 0$ and $s_i > 3\delta_i$ and $s_i' = 3\delta_i$; the other cases are treated similarly. Now if $f'(x_i) \le 3\delta_i$, then s_i' is a better approximation to $f'(x_i)$ than s_i and therefore (8.13) follows. On the other hand, if $f'(x_i) = 3\delta_i + \eta$ with $\eta > 0$ then we let l be the linear function which interpolates f' at x_i and x_{i+1} if $x_i < 1$ and at x_{i-1} and x_i if $x_i = 1$. We shall assume that $x_i < 1$; the case $x_i = 1$ is handled similarly. Since $f'(x_{i+1}) \ge 0$, we have

$$\int_{x_i}^{x_{i+1}} (l(x) - f'(x))\,dx \ge \Delta x_i\,(3\delta_i + \eta)/2 - \Delta x_i \delta_i \ge \Delta x_i \eta/2\,. \tag{8.14}$$

On the other hand by Newton's formula (7.3) of Chapter 4, we have

$$|f'(x) - l(x)| = |[x_i, x_{i+1}, x]\,f'|\,(x - x_i)\,(x_{i+1} - x) \le \varepsilon_i/8\,, \quad x \in [x_i, x_{i+1}]\,.$$

Therefore the integral on the left side of (8.14) does not exceed $\varepsilon_i(\Delta x_i)/8$. We conclude that $\eta \le \varepsilon_i/4$. This yields (8.13) with s_i replaced by its new value s_i'.

Now let Q_i be the Hermite cubic interpolant (3.4) to f on $[x_i, x_{i+1}]$. From (3.3), we have

$$\begin{aligned} |P_i(x) - Q_i(x)| &\le C\Delta x_i\,[|s_i - f'(x_i)| + |s_{i+1} - f'(x_{i+1})|] \\ &\le C\,[\varepsilon_i \Delta x_i + \varepsilon_{i+1}\Delta x_{i+1}]\,. \end{aligned}$$

Therefore, (8.12) follows from (3.6) and the estimate (see (5.7) of Chapter 2)

$$E_3(f', [x_i, x_{i+1}]) \le E_2(f', [x_i, x_{i+1}]) \le \|f^{(3)}\| h^2\,. \qquad \square$$

§ 9. Shape Preserving Quadratic Spline Interpolation

There is another approach to shape preserving interpolation, this time by quadratic splines. At the given points x_i, $0 = x_0 < x_1 < \dots < x_n = 1$, we interpolate the data (y_i), (which we will always assume to be produced by a function f). We want the spline interpolant $S \in C^1$ to mimic the shape of the data, their monotonicity $(\Delta y_i := y_{i+1} - y_i \ge 0)$ or convexity. The data (y_i) is convex (or concave) if $\delta_i := \Delta y_i/\Delta x_i$ satisfy $\Delta\delta_i \ge 0$ (or $\Delta\delta_i \le 0$), $i = 0, \dots, n-2$. The spline space $\mathcal{S} := \mathcal{S}_3(T, [0,1])$ from which we take our interpolants will consist of quadratic splines with simple knots at the interior interpolation points x_i, $i = 1, \dots, n-1$, and at additional points ξ_i, $x_i < \xi_i < x_{i+1}$, $i = 0, \dots, n-1$. Thus $\dim(\mathcal{S}) = 2n + 2$. We interpolate according to the rule

$$S(x_i) = y_i\,, \quad S'(x_i) = s_i\,, \quad i = 0, \dots, n\,. \tag{9.1}$$

Here, the slopes s_i are not necessarily the values $f'(x_i)$ but free parameters. The s_i and ξ_i are selected in a way so that the interpolant S will inherit the shape of the data.

The Hermite interpolation problem (9.1) satisfies the interlacing condition (9.6) of Chapter 5 and by the Karlin-Ziegler theorem has a unique solution.

We first describe when a quadratic spline $S \in \mathcal{S}$ is monotone or convex on $[x_i, x_{i+1}]$. For this we compute $S'(\xi_i)$. Since the integral of the piecewise linear function S' over $[x_i, x_{i+1}]$ must equal Δy_i, we have

$$\begin{gathered} S'(\xi_i) = 2\delta_i - \alpha_i s_i - \beta_i s_{i+1}\,, \\ \alpha_i := (\xi_i - x_i)/\Delta x_i \text{ and } \beta_i := (x_{i+1} - \xi_i)/\Delta x_i\,. \end{gathered} \tag{9.2}$$

The right side of (9.2) is $\mu_i(S)$ where μ_i is the linear functional

$$\mu_i(f) := 2\delta_i(f) - \alpha_i f'(x_i) - \beta_i f'\,(x_{i+1}) \tag{9.3}$$

with $\delta_i(f) := (f(x_{i+1}) - f(x_i))\,/\,(x_{i+1} - x_i)$.

Lemma 9.1 (Schumaker [1983]). *A spline $S \in \mathcal{S}$ will be nondecreasing (nonincreasing) on $[x_i, x_{i+1}]$ if and only if*

$$s_i\,, \quad s_{i+1}\,, \text{ and } 2\delta_i - \alpha_i s_i - \beta_i s_{i+1} \text{ are all } \geq 0 \ (\leq 0)\,. \tag{9.4}$$

Furthermore, S will be convex on (x_i, x_{i+1}) if and only if

$$s_i \leq 2\delta_i - \alpha_i s_i - \beta_i s_{i+1} \leq s_{i+1} \tag{9.5}$$

and S will be concave on this interval if and only if

$$s_i \geq 2\delta_i - \alpha_i s_i - \beta_i s_{i+1} \geq s_{i+1}\,. \tag{9.6}$$

Proof. The derivative S' is a continuous piecewise linear function with breakpoints at the x_j, $j = 1, \ldots, n-1$, and the ξ_j, $j = 0, \ldots, n-1$. Since $S'(\xi_i) = 2\delta_i - \alpha_i s_i - \beta_i s_{i+1}$, S' will be nonnegative if and only if all terms of (9.4) are ≥ 0. Similarly for nonpositive S'. Since S is convex if and only if S' is increasing we also have (9.5); similarly for (9.6). □

Given nonnegative values s_i, s_{i+1} and δ_i, there may not be any choice of ξ_i in (x_i, x_{i+1}) for which (9.4) holds. For example if $\delta_i = 0$, then there is a choice if and only if $s_i = s_{i+1} = 0$. If $\delta_i \neq 0$, there is a choice if and only if one of the two numbers s_i, $s_{i+1} < 2\delta_i$ or $s_i = s_{i+1} = 2\delta_i$. A similar statement holds when $\delta_i < 0$. The two cases of (9.4) can therefore be combined into a single statement. Namely, there is a choice $\xi_i \in (x_i, x_{i+1})$ such that S is monotone on $[x_i, x_{i+1}]$ if and only if

$$0 \leq \min\left(s_i/\delta_i, s_{i+1}/\delta_i\right) < 2 \text{ or } s_i = s_{i+1} = 2\delta_i\,. \tag{9.7}$$

When (9.7) is valid, the set of points ξ_i for which (9.7) holds is a nontrivial interval which we call the *admissible interval for monotonicity.* This interval contains points near x_{i+1} if s_i/δ_i is less than 2, similarly for x_i. In the case $s_i = s_{i+1} = 2\delta_i$, this interval is (x_i, x_{i+1}).

A similar situation holds for convexity and concavity. Given s_i, s_{i+1}, and δ_i, the set of ξ_i for which either of (9.5) or (9.6) is satisfied is a (possibly empty) subinterval of (x_i, x_{i+1}) called the *admissible interval for convexity.* This interval is nonempty if

$$s_i < \delta_i < s_{i+1} \text{ or } s_i > \delta_i > s_{i+1} \text{ or } s_i = s_{i+1} = \delta_i\,. \tag{9.8}$$

For example, in order that S be convex on $[x_i, x_{i+1}]$, (9.5) must hold or equivalently

$$(1 + \alpha_i)s_i + \beta_i s_{i+1} \leq 2\delta_i \leq (1 + \beta_i)s_{i+1} + \alpha_i s_i\,.$$

Therefore, if $s_i = s_{i+1}$, then δ_i must share this value. While if $s_i \neq s_{i+1}$ then since α_i, $\beta_i > 0$, $\alpha_i + \beta_i = 1$, we must have the first statement of (9.8). Similarly S is concave on $[x, x_{i+1}]$ if and only if one of the last two statements of (9.8) holds. The admissible interval for convexity is always a subinterval of the admissible interval for monotonicity when the latter is nonempty. In fact, if ξ is a point in the admissible interval for convexity and if s_i and s_{i+1} are not of opposite sign then by (9.5) and (9.6), $S'(\xi)$ shares their sign. Hence the piecewise linear function S' does not change sign on $[x_i, x_{i+1}]$ and therefore S is monotone on this interval.

The algorithms for generating monotonicity and convexity preserving quadratic splines rest on how the slopes s_i and the knots ξ_i are chosen. We mention two examples. We let

$$H_i := \frac{2\delta_{i-1}\delta_i}{\delta_{i-1} + \delta_i} \tag{9.9}$$

be the *harmonic mean* of δ_{i-1} and δ_i. The first algorithm defines the slopes at the interior x_i as follows:

$$s_i := \begin{cases} 0, & \text{if } \delta_{i-1}\delta_i \leq 0 \\ H_i\,, & \text{otherwise} \end{cases}, \quad i = 1, \ldots, n-1\,. \tag{9.10}$$

This formula cannot be used at the endpoints since the apropriate values of δ_i are not available. Instead, we define the endpoint slopes as follows:

$$\begin{gathered} s_0 := \begin{cases} 0\,, & \text{if } \delta_0\,(2\delta_0 - s_1) \leq 0 \\ 2\delta_0 - s_1\,, & \text{otherwise} \end{cases} \\ s_n := \begin{cases} 0, & \text{if } \delta_{n-1}\,(2\delta_{n-1} - s_{n-1}) \leq 0, \\ 2\delta_{n-1} - s_{n-1}, & \text{otherwise.} \end{cases} \end{gathered} \tag{9.11}$$

With this choice of slopes, it is easy to see that the admissible interval for monotonicity is always the entire interval (x_i, x_{i+1}). Indeed, s_i and s_{i+1} are never of opposite signs and therefore (9.4) holds because $|s_i|$, $|s_{i+1}| \leq 2\delta_i$.

Now we can describe our interpolant S. Given (y_i), we let ξ_i be a point from the interval for convexity in (x_i, x_{i+1}) if this interval is nonempty. Otherwise, we let ξ_i be a point from the interval for monotonicity which as we have just observed is always nonempty. We then define S as the spline in $\mathcal{S}_3(T, I)$ which satisfies the interpolation conditions (9.1) for the slope assignment (9.10), (9.11). Then S is comonotone with the data (y_i). That is, if $y_i \le y_{i+1} \le \cdots \le y_j$, then S is nondecreasing on $[x_i, x_j]$. Similarly, if the data is decreasing.

We can show that if $\Delta\delta_i > 0$, $i = 0, \dots, n-2$, then S is *convex* on I. Indeed, from the definition (9.10), we have $\delta_{i-1} \le s_i \le \delta_i$, $i = 1, \dots, n-1$. Hence $\delta_0 \le s_1 \le \delta_1 \le \cdots \le s_{n-1} \le \delta_{n-1}$. Since $s_1 \ge \delta_0$, we also have $2\delta_0 - s_1 \le \delta_0$. Therefore, the definition (9.11) of s_0 gives that $s_0 \le \delta_0$. Similarly, we have $\delta_{n-1} \le s_n$ and therefore

$$s_0 \le \delta_0 \le s_1 \le \cdots \le \delta_{n-1} \le s_n \,. \tag{9.12}$$

Referring again to the definitions (9.10), (9.11), we see that an equality $s_j = \delta_j$ or $s_{j+1} = \delta_j$ can hold in (9.12) if and only if $\delta_j = 0$ and in this case $s_j = s_{j+1} = \delta_j$. Therefore (9.8) is satisfied. This shows that S is convex on each interval $[x_i, x_{i+1}]$ and hence on I.

A special case of the above algorithm was given by McAllister and Roulier [1981] (who describe their interpolant, however, by means of Bernstein polynomials). Their algorithm corresponds to choosing the knots ξ_i as the midpoint of the convexity interval if this interval is nonempty and as the midpoint of the monotonicity interval otherwise.

One disadvantage of the above algorithms is that they do not in general give the best order of approximation possible by splines in $\mathcal{S}_3(T, I)$. For example, the function x^2 is only approximated with order $\mathcal{O}(h^2)$ with $h := \max \Delta x_i$. This limitation on the order of approximation reflects the fact that H_i is only a first order approximation to $f'(x_i)$ if the latter is small.

The next algorithm we describe makes a change in the slope assignments (9.10), (9.11) in order to improve the rate of approximation. From the numerical differentiation formula (8.6), it follows that

$$\sigma_i := (\delta_{i-1}\Delta x_i + \delta_i \Delta x_{i-1}) \,/\, (\Delta x_{i-1} + \Delta x_i)\,, \quad i = 1, \dots, n-1\,, \tag{9.13}$$

is a *second order* approximation to $f'(x_i)$. We follow DeVore and Yan [1986] and modify the slope assignments (9.10) by using σ_i in place of H_i when σ_i is small: for $i = 1, \dots, n-2$,

$$s_i := \begin{cases} 0\,, & \text{if } \delta_{i-1}\delta_i \le 0\,, \\ H_i\,, & \text{if } \delta_{i-1}\delta_i > 0 \text{ and } \min(\sigma_i/\delta_i, \sigma_{i+1}/\delta_i) \ge 2\,, \\ \sigma_i\,, & \text{otherwise}\,. \end{cases} \tag{9.14}$$

The definition (9.14) does not apply when $i = n-1$ because σ_n is not defined; in this case, we use

$$s_{n-1} := \begin{cases} 0\,, & \text{if } \delta_{n-2}\delta_{n-1} \le 0\,, \\ \sigma_{n-1}\,, & \text{otherwise}\,. \end{cases} \tag{9.15}$$

The slope assignments for s_0 and s_n are the *same* as in (9.11).

With the above slope assignments, we choose ξ_i as a point from the admissible interval for convexity if this interval is nonempty and otherwise as a point from the admissible interval for monotonicity which we now show is always nonempty.

Lemma 9.2. *For arbitrary data (y_i) and the above slope assignments, the admissible intervals for monotonicity are nonempty for $i = 0, \dots, n$.*

Proof. If $\delta_i = 0$, then $s_i = s_{i+1} = 0$ and therefore the admissible interval for monotonicity is all of (x_i, x_{i+1}). If $\delta_i \ne 0$, it is enough to show that (9.7) holds and therefore we may also assume that $s_i \ne 0$, $s_{i+1} \ne 0$. We consider the remaining possibilities.

Case $i = 0$. It follows from (9.11) that δ_0 and $2\delta_0 - s_1$ are of the same sign. If they are both positive, then $s_0 = 2\delta_0 - s_1$ and hence (9.7) holds; similarly when both are negative.

Case $i = n-1$. This is similar to the above case.

Case $0 < i < n-1$. Since s_i, $s_{i+1} \ne 0$, it follows from (9.14) that δ_{i-1}, δ_i, and δ_{i+1} are all of the same sign and s_i, s_{i+1} share their sign. We assume that they are positive; a similar argument applies when they are negative. If s_i is defined to be H_i by (9.14) then $s_i < 2\delta_i$; similarly if $s_{i+1} := H_{i+1}$. This gives (9.7). In the remaining case $s_i := \sigma_i$ and $s_{i+1} := \sigma_{i+1}$ and therefore the criteria in (9.14) for defining s_i give (9.7). □

Theorem 9.3 (DeVore-Yan [1986]). *The spline $S \in \mathcal{S}_3(T, I)$ which satisfies the interpolation conditions* (9.1) *for the slope assignments* (9.14), (9.15), *and* (9.11) *is comonotone with the data (y_i). Also, if $\Delta\delta_i > 0$, $i = 0, \dots, n-2$, then S is convex. If f is a monotone function in $C^3(I)$, then*

$$\|f - S\| \le 3 \left\| f^{(3)} \right\| h^3\,, \quad h := \max \Delta x_i\,. \tag{9.16}$$

Proof. The monotonicity and convexity preserving properties of S are proved in the same way as for the first algorithm of this section. To prove (9.16), we assume that f is nondecreasing and estimate $|f'(x_i) - s_i|$. Let $M := \|f^{(3)}\|$. Calculating the divided difference of f, we find from (7.4) of Chapter 4,

$$|f'(x_i) - \sigma_i| = |\Delta x_{i-1}\Delta x_i\,[x_{i-1}, x_i, x_i, x_{i+1}]\,f| \le Mh^2/6\,, \quad 0 < i < n\,. \tag{9.17}$$

A similar inequality holds for $|f'(x_i) - H_i|$, $i = 1, \dots, n-2$, provided $\min(\sigma_i, \sigma_{i+1}) \ge 2\delta_i$. Indeed, σ_i and H_i both lie between δ_{i-1} and δ_i and therefore by (9.17), we have

$$|f'(x_i) - H_i| \le |f'(x_i) - \sigma_i| + |\sigma_i - H_i| \le Mh^2/6 + |\delta_i - \delta_{i-1}|\,. \tag{9.18}$$

To estimate $\delta_i - \delta_{i-1}$, we use the following divided difference

$$[x_{i-1}, x_i, x_{i+1}, x_{i+2}]\,f = \frac{\frac{\delta_{i+1} - \delta_i}{x_{i+2} - x_i} - \frac{\delta_i - \delta_{i-1}}{x_{i+1} - x_{i-1}}}{x_{i+2} - x_{i-1}}\,. \tag{9.19}$$

Since σ_i is a convex combination of δ_{i-1} and δ_i and by our assumption $\sigma_i \ge 2\delta_i$, we have $\delta_{i-1} \ge 2\delta_i$. Similarly, $\delta_{i+1} \ge 2\delta_i$. Therefore the numerator on the right side of (9.19) is the sum of two non-negative terms and hence each of these terms is smaller than $(x_{i+2} - x_{i-1})\, M/6$. This gives $|\delta_i - \delta_{i-1}| \le Mh^2$, $i = 1, \dots, n-1$. Using this in (9.18) shows that

$$(9.20) \qquad |f'(x_i) - H_i| \le 2Mh^2\,, \qquad 0 < i \le n-2\,,$$

provided $\min(\sigma_i, \sigma_{i+1}) \ge 2\delta_i$. We can now prove that

$$(9.21) \qquad |f'(x_i) - s_i| \le \begin{cases} 2Mh^2\,, & 0 < i < n \\ 3Mh^2\,, & i = 0, n\,. \end{cases}$$

Consider first the case $0 < i < n$. If $s_i = H_i$ or $s_i = \sigma_i$, then this is (9.17) or (9.20) respectively. If $s_i = 0$, then either $\delta_{i-1} = 0$ or $\delta_i = 0$ and hence f' vanishes identically on an interval which contains x_i and (9.21) follows.

We now check the case $i = 0$; the case $i = n$ is similar. We have

$$(9.22) \qquad |2\delta_0 - f'(x_0) - f'(x_1)| = (x_1 - x_0)^2\, |[x_0, x_0, x_1, x_1]\, f| \le Mh^2/6\,.$$

Using this with (9.21) for $i = 1$, we have that $|2\delta_0 - s_1 - f'(x_0)| \le 3Mh^2$. This is the desired inequality when s_0 is defined to be $2\delta_0 - s_1$. On the other hand, if $s_0 := 0$, then by the definition of s_0, we have two possibilities. If $\delta_0 \le 0$, then $\delta_0 = 0$ because f is increasing; hence $f'(x_0) = 0$ and so (9.21) is obvious. If $2\delta_0 - s_1 \le 0$, then by (9.22) and the first part of (9.21), we find that $f'(x_0) \le 3Mh^2$ and (9.21) follows in this case as well.

Now the linear functional μ_i of (9.3) differentiates quadratic polynomials Q exactly: $\mu_i(Q) = Q'(\xi_i)$. If Q is the quadratic Taylor polynomial of f at ξ_i, then $\|f' - Q'\|\, [x_i, x_{i+1}] \le Mh^2/2$. Hence,

$$\begin{aligned} |f'(\xi_i) - \mu_i(f)| &= |\mu_i(f - Q)| \\ &\le (2 + |\alpha_i| + |\beta_i|)\, \|f' - Q'\|\ [x_i, x_{i+1}] \le 3Mh^2/2\,. \end{aligned}$$

Using this with (9.21) and the fact that $S'(\xi_i)$ is given by (9.2), we find

$$|f'(\xi_i) - S'(\xi_i)| \le 5Mh^2\,.$$

This shows that the piecewise linear function S' approximates f' to an error at most $5Mh^2$ at each of its breakpoints. Hence if T is the piecewise linear interpolant to f' at these breakpoints, we have $\|T - S'\| \le 5Mh^2$. Also, $\|f' - T'\| \le Mh^2/8$. Indeed, if $x_i \le x \le x_{i+1}$, then $|f'(x) - T'(x)| = |(x - x_i)(x - x_{i+1})\, [x_i, x_{i+1}, x]\, f| \le Mh^2/8$. This gives $\|f' - S'\| \le 6Mh^2$. Since $f - S$ vanishes at each x_i, $i = 0, \dots, n$, we obtain (9.16) by integrating from (the nearest) x_i to x. □

§ 10. Problems

10.1. If $f \in C^2[0,1]$ is monotone on $[0,1]$, then the quadratic spline interpolant S to f at points x_i, $i = 1, \dots, n$, of the McAllister-Roulier algorithm satisfies $\|f - S\|_\infty \le C \left\| f^{(2)} \right\| h^2$, $h := \max x_{i+1} - x_i$. For $f(x) = x^2$, $\|f - S\|_\infty \ge Ch^2$.

10.2. (Ciesielski [1963]) For $r = 2$ there is a constant $C > 0$ such that for any knot sequence T, the L_2 spline projection has norm $\le C$ on $L_\infty(\mathbb{R})$.

10.3. If in addition to (7.9), $D^j\widehat{\varphi}(0) = 0$, $j = 1, \dots, r-1$, then $P_j(x) = x^j$, $j = 0, \dots, r-1$, for the polynomials P_j of Theorem 7.1.

10.4. If for a continuous function φ of compact support $\sum_{k\in\mathbb{Z}} c_k \varphi(x - k) \equiv 1$ with the c_k uniformly bounded, then $(\gamma\varphi(\cdot - k))$ is a partition of unity for some real number γ.

10.5. If $r > 2$, there are points $X = (x_j)_{-r+1}^{n}$ such that for all knot sequences $T = (t_j)_{-r+1}^{n+r}$, the x_i cannot be the knot averages (1.15) of T.

10.6. (Eisenstadt-Jackson-Lewis [1985]) If the algorithm (8.5) assigns slopes s_i so that $(s_i, s_{i+1}) \in \delta_i M'$ where M' is a fixed subregion of M and if for each monotone f in $C^3[0,1]$, the spline S of this algorithm (where $y_i := f(x_i)$) approximates f to order $\mathcal{O}(h^3)$, $h := \max \Delta x_i$, then M' contains the triangle with vertices $(0,0)$, $(2,0)$ and $(0,2)$.

§ 11. Notes

11.1. Marsden's proof [1974] of Theorem 2.1 shows that $\|L(X,T)\| \le 2$. The proof we have given using the total positivity of the collocation matrix will give $\|\mathcal{S}(X,T)\| \le 8$ with a more careful choice of the sequence c (de Boor [1975]). Total positivity can also be used to show that cubic spline interpolation at the knot averages (1.15) gives a projection $L(X,T)$ whose norm can be bounded independently of the knot sequence T (de Boor [1975]). However, Jia, in 1988 has shown that for large degree and interpolation at knot averages $\|L(X,T)\|$ can be made arbitrarily large by varying the knots T.

11.2. The matrices which occur in describing the L_2 spline projection are also totally positive. This can be shown by discretizing the integrals $\int_{-\infty}^{\infty} N_i N_j\, dx$ and using the total positivity of the collocation matrix. This gives another approach to estimating the norm of the L_2 spline projection on C_B. For example, de Boor [1976$_2$] uses this to show that for $r = 3$ or 4, the norm of the L_2 spline projection is bounded independently of the knots T. Another application of total positivity is to derive boundedness of the L_2 spline projection on a finite interval from a corresponding result for $\mathbb{R}$. This uses the fact (de Boor-Jia-Pinkus [1982]) that a boundedly invertible totally positive bi-infinite matrix A has principal submatrices whose inverses have norms not exceeding $\|A^{-1}\|$.

11.3. The results of this chapter which estimate the norm of the bi-infinite interpolation projection or the L_2 spline projection require that the knots have bounded mesh ratios. There is therefore special interest in knot sequences which do not have bounded mesh ratios; in particular in geometric knot sequences $T := (q^\nu)_{\nu\in\mathbb{Z}}$, $q > 1$, and the corresponding spline spaces $\mathcal{S}_r(T, \mathbb{R}_+)$.

Micchelli [1976] has shown that there are exactly $r - 1$ values of q for which spline interpolation at the knots T is not a bounded projection on $C(\mathbb{R}_+)$. Höllig and Scherer [1980] have shown that there is a uniform (in terms of q) bound on the

L_∞-norm of the L_2-spline projection. A unified approach to these questions using the block Toeplitz structure of the corresponding matrices was given by Höllig [1981].

11.4. From theorems which bound the L_2 spline projection on $\mathbb{R}$, we can derive corresponding theorems for $\mathcal{S}_r(T, I)$, $I := [0,1]$. For example, let $t_1, \dots, t_n$ be the basic knots of T and $t_i := 0$, $i = -r+1, \dots, 0$, $t_i := 1$, $i = n+1, \dots, n+r$. If $t_{i+r} - t_i \le M\,(t_{j+r} - t_j)$, $i, j = -r+1, \dots, n$, then we can extend T to a bi-infinite knot sequence T' which satisfies (4.2). If $f \in C(I)$, we let $\tilde{f}$ be an extension of f to $\mathbb{R}$ which is zero outside $[-2,2]$ and satisfies $\|\tilde{f}\|_\infty \le \|f\|_\infty$. If S is the best $L_2(\mathbb{R})$ approximation to $\tilde{f}$ from $\mathcal{S}_r(T', \mathbb{R})$, then the restriction of S to I is the best L_2 approximation to f from $\mathcal{S}_r(T, I)$. Indeed, the only B-splines N_j which are non-zero on I are supported entirely in I and therefore the orthogonality conditions

$$\int_0^1 (f - S)N_j dx = \int_{-\infty}^{\infty} (\tilde{f} - S)N_j dx = 0, \quad j = -r+1, \dots, n$$

are valid. From Theorem 4.4, we have $\|P_T\|_{C(I)} \le \|P_{T'}\|_{L_\infty} \le C\sqrt{M}$. A similar conclusion can be drawn for $1 \le p < \infty$.

11.5. An orthogonal wavelet is a function φ whose normalized translated dilates $\varphi_{j,k}(x) := 2^{k/2}\varphi(2^k x - j)$, $k \in \mathbb{Z}$, $j \in \mathbb{Z}^d$, are an orthonormal basis for $L_2(\mathbb{R})$:

$$f = \sum_{j,k\in\mathbb{Z}} \langle f, \varphi_{j,k}\rangle \varphi_{j,k}\,, \qquad f \in L_2(\mathbb{R})\,.$$

A simple example is the Haar orthogonal system which is generated by the step function $\varphi(x) := 1$, $x \in [0, \frac{1}{2})$, $= -1$, $x \in [\frac{1}{2}, 1]$, respectively, and the value zero otherwise. Meyer (see [B-1990]) and Daubechies [1988] have constructed orthogonal wavelets of arbitrary smoothness C^r which have compact support and form a basis also for L_p, Hardy spaces H_p, Sobolev and Besov spaces. Moreover, the norms in these spaces can often be described simply in terms of the coefficients $< f, \varphi_{j,k} >$. For example, if $1 < p \le \infty$, $0 < q \le \infty$, $\alpha > 0$, then the expression (3.15) of §3, Chapter 12, with $a_{j,n}$ replaced by $< f, \varphi_{j,n} >$ is an equivalent semi-norm for the Besov spaces $B_q^\alpha(L_p)$.

11.6. If φ is a continuous function with compact support on $\mathbb{R}$, then functions $S = \sum c_{j,k}\varphi(2^k x - j)$ with at most n nonzero coefficients form a nonlinear manifold $\mathcal{M}_n$. For example, if φ is the B-spline $M(x; 0, 1, \dots, r)$ of order r, then $\mathcal{M}_n$ is a submanifold of $\Sigma_{rn,r}$. In this case, DeVore, Jawerth and Popov [1991] show that Theorems 8.3 and 8.4 of Chapter 12 remain valid for approximation by $\mathcal{M}_n$, $n = 1, 2, \dots$. Similar theorems hold for other functions φ of one or several variables even without compact support. For example, if $\varphi(x) := (1 + x^2)^{-m}$, then $\mathcal{M}_n$ is a manifold of rational functions of degree n. Theorems 8.3 and 8.4 remain true for approximation in L_p, $1 < p < \infty$ by the elements of $\mathcal{M}_n$ provided m is sufficiently large.

11.7. The results of §7 have two important generalizations. In the first, the function φ be can be replaced by a family of functions $\{\varphi_j\}_1^m$. The second (Strang-Fix [1973]) extends Lemma 7.1 to functions of several variables. This is particularly useful in the study of multivariate spline approximation.

11.8. De Boor and Jia [1985] have given a certain converse to Theorem 7.4 for compactly supported functions φ of the type studied in §7. If there is a constant $\delta > 0$ such that each f in the unit ball of $W_p^r(\mathbb{R})$ can be approximated by $S_h(x) := \sum_{k\in\mathbb{Z}} a_k(h)\phi\left(\frac{x}{h} - k\right)$ with error $\mathcal{O}(h^r)$ and $a_k(h) = 0$ whenever $\operatorname{dist}(kh, \operatorname{supp}(f)) > \delta$ then $\mathcal{S}(\phi)$ is said to provide *local approximation of order* r. De Boor and Jia prove

that whenever $\mathcal{S}(\varphi)$ provides local approximation of order r, the function φ must satisfy (7.10).

11.9. Eisenstadt-Jackson-Lewis [1985] have shown that for monotone functions $f \in C^4[0,1]$, there is a selection of slopes s_i, $i = 0, \dots, n$, so that algorithm (8.5) yields a monotone cubic spline which approximates f with error $\le C\left\|f^{(4)}\right\| h^4$. In step (i), the slopes are chosen as third order approximations to $f'(x_i)$. At step (iii) they are modified so that $(s_i, s_{i+1}) \in \delta_i M$. This modification is fairly intricate as it involves making both a forward sweep and then a backward sweep through the s_i with possible changes of each s_i in each sweep. Another approach given by Yan [1987] to obtaining fourth order monotonicity preserving cubic interpolants is to insert two additional knots.

Bibliography

A. Books on Approximation

Quoted in the text as, for example, Akhiezer [A-1956]

Akhiezer, J. (1956) Theory of approximation. Ungar, New York (Translation from the Russian). There exists an augmented 2nd edition, Nauka, Moscow, 1965 and its translation into German
Bernstein, S. (1926) Leçons sur les propriétés extrémales et la meilleure approximation des fonctions analytiques d'une variable réelle. Gauthier-Villars, Paris
Bernstein, S.N. (1952, 1954) Collected works, vols. I and II (Russian). Akad. Nauk. SSSR, Moscow
Besikovich, Ya.S. (1939) Calculus of finite differences. Leningrad University Press
Bojanov, B.D., Hakopian, H.A. and Sahakian, A.A. (1993) Spline functions and multivariate interpolation. Kluwer Acad. Publ., Dodrecht
de Boor, C. (1978) A practical guide to splines. Springer, New York
de Boor, C. (1990) Splinefunktionen. Birkhäuser, Basel
Butzer, P.L. and Berens, H. (1967) Semi-groups of operators and approximation. Springer, Berlin
Butzer, P.L. and Nessel, R.J. (1971) Fourier analysis and approximation. Academic Press, New York
DeVore, R.A. (1972) The approximation of continuous functions by positive linear operators. Lecture Notes in Mathematics, vol. 293. Springer, Berlin
Ditzian, Z. and Totik, V. (1987) Moduli of smoothness. Springer, New York
Dzyadyk, V.K. (1977) Introduction to the theory of uniform approximation of functions by polynomials. Nauka, Moscow
Feinerman, R.P. and Newman, D.J. (1974) Polynomial approximation. Williams and Wilkins, Baltimore
Gaier, D. (1980) Vorlesungen über Approximation im Komplexen. Birkhäuser, Basel
Jackson, D. (1930) The theory of approximations. Amer. Math. Soc., New York
Karlin, S. (1968) Total positivity. Stanford University Press
Karlin, S. and Studden, W.J. (1966) Tchebycheff systems: with applications to analysis and statistics. Interscience, New York
Korneichuk, N.P. (1976) Extremal problems of approximation theory. Nauka, Moscow (Russian)
Korneichuk, N.P. (1984) Splines in approximation theory. Nauka, Moscow (Russian)
Korneichuk, N.P. (1987) Exact constants in approximation theory. Nauka, Moscow (Russian); English edition: Cambridge Univ. Press, 1991
Korovkin, P.P. (1959) Linear operators and approximation theory. Fizmatgiz, Moscow; English edition: Hindustan Publ. Corp., Delhi, 1960
Lorentz, G.G. (1966, 1986) Approximation of functions. 1st edition: Holt, Rinehart and Winston, New York 1966; 2nd edition: Chelsea, New York 1986
Lorentz, G.G. (1953) Bernstein polynomials. 1st edition: Univ. of Toronto Press, Toronto 1953; 2nd edition: Chelsea, New York 1986
Lorentz, G.G., Jetter, K. and Riemenschneider, S.D. (1983) Birkhoff interpolation. Encyclopedia of Mathematics and its Applications, vol. 19. Addison-Wesley, Reading 1983
Lorentz, R.A. (1992) Multivariate Birkhoff interpolation. Lecture Notes in Mathematics, vol. 1516. Springer, Berlin
Malozemov, V.N. (1973) Simultaneous approximation of a function and its derivatives (Russian). Len. Gos. Univ., Leningrad
Natanson, I.P. (1964–1965) Constructive function theory, 3 vols. Ungar, New York (Translation from Russian)
Nürnberger, G. (1989) Approximation by spline functions. Springer, Berlin
Obreschkov, N. (1963) Verteilung und Berechnung der Nullstellen reeller Polynome. Deutscher Verlag der Wissenschaften, Berlin
Pinkus, A. (1989) On L^1-approximation. Cambridge Tracts, vol. 93. Cambridge University Press
Powell, M.J.D. (1981) Approximation theory and methods. Cambridge University Press, Cambridge
Rahman, Q.A. and Schmeisser, G. (1983) Les inégalités de Markoff et de Bernstein. Univ. Montreal Press, Montreal
Rice, J.R. (1964 and 1969) The approximation of functions, vols.1 and 2. Addison-Wesley, Reading
Rivlin, T.J. (1990) The Chebyshev polynomials, 2nd edn. Wiley and Sons, New York
Schoenberg, I.J. (1973) Cardinal spline interpolation. Regional Conference Series in Appl. Math. vol. 12. SIAM, Philadelphia
Schumaker, L.L. (1981) Spline functions: basic theory. Wiley and Sons, New York
Szabados, J. and Vértesi, P. (1990) Interpolation of functions. World Scientific, Singapore
Timan, A.F. (1963) Theory of approximation of functions of a real variable. MacMillan, New York
Tureckii, A.Ch. (1968 and 1977) Theory of interpolation in problems, vols. 1 and 2. Izd. Vyssh. Shkola, Minsk (Russian)
de la Vallée-Poussin, C. (1970) Leçons sur l'approximation des fonctions d'une variable réelle. Gauthier-Villars, Paris 1919; 2nd edn. Chelsea Publ. Co., New York 1970
Zielke, R. (1979) Discontinuous Cebysev systems. Lecture Notes in Mathematics, vol. 707. Springer, Berlin

B. Other Books

Quoted in the text as, for example, Adams [B-1975]

Adams, R.A. (1975) Sobolev spaces. Academic Press, New York
Bennett, C. and Sharpley, R. (1988) Interpolation of operators. Academic Press, New York
Bergh, J. and Löfström, J. (1976) Interpolation spaces. Springer, Berlin
Chebyshev, P.L. (1962) Oevres, vol. 1. Chelsea Publ. Co., New York, pp. 273–378
Dunford, N. and Schwartz, J.T. (1958) Linear operators, part I. Interscience Publ., New York
Duren, P.L. (1970) Theory of H^p spaces. Academic Press, New York
Eggleston, H.G. (1958) Convexity. Cambridge Tracts, vol. 27. Cambridge University Press
Farin, G.F. (1988) Curves and surfaces for computer aided geometric design. A practical guide. Academic Press, New York
Freud, G. (1969) Orthogonale Polynome. Birkhäuser, Basel
Halmos, P.R. (1950) Measure theory. Van Nostrand, New York

Hardy, G.H. (1949) Divergent series. Clarendon Press, Oxford
Hardy, G.H., Littlewood, J.E. and Pólya, G. (1964) Inequalities, 2nd edn. Cambridge University Press
Hille, E. and Phillips, R.S. (1957) Functional analysis and semi-groups. Amer. Math. Soc., Providence
Hille, E. (1962) Analytic function theory, vols. 1, 2. Gin and Co., Boston
Johnson, A., Wallin, H. (1984) Functions spaces on subsets of R^n. Reports 2, part 1. Harwood Academic Publishers, London
Katznelson, Y. (1968) An introduction to harmonic analysis. Wiley and Sons, New York
Landau, E. and Gaier, D. (1986) Darstellung und Begründung einiger neuerer Ergebnisse der Funktionentheorie, 3rd edn. Springer, Berlin
Meyer, Y. (1990) Ondelettes et operateurs, vol. 1. Herman, Paris
Natanson, I.P. (1960) Theory of functions of a real variable, vol. II. Ungar, New York
Peetre, J. (1963) A theory of interpolation of normed spaces. Notes Universidade de Brasilia
Royden, H.L. (1968) Real analysis, 2nd edn. MacMillan, New York
Strang, (1976) Linear algebra and its applications. Academic Press, New York
Szegö, G. (1975) Orthogonal polynomials, 4th edn. Amer. Math. Soc., Providence
Yosida, K. (1965) Functional analysis. Springer, Berlin

C. Articles

Quoted in text as, for example, Alexits [1941]; Ch. 11:341 is the reference to the chapter and the page of the book where it is quoted

Abbreviations: JAT = J. Approx. Theory; CA = Const. Approximation; BAMS or PAMS or TAMS = Bulletin, or Proceedings, or Translations of the American Mathematical Society; Dokl. = Dokl. Akad. Nauk SSSR; Izv. = Izv. Akad. Nauk SSSR, Ser. Mat.; SIAM Num. = SIAM J. Num. Anal.; Mat. Sb. = Matem. Sbornik, N.S.

Akhiezer, N.I., Krein, M.G. (1937) On the best approximation of periodic functions. Dokl. **15**, 107–112; Ch. 7:237
Alexits, G. (1941) Sur l'ordre de grandeur de l'approximation d'une fonction par les moyennes de sa série de Fourier. (Hungar.) Mat. Fiz. Lapok **48**, 410–422; Ch. 11:341
Amelkovic, V.C. (1966) A theorem converse to a theorem of Voronovskaya type. Teor. Funkcii Anal. i Prilozhen. 67–74; Ch. 10:331
Arama, A.O. (1960) Properties concerning the monotonicity of the polynomials (in Russian). Mathem. Cluj **2** (25) 25–40; Ch. 10:310
Arestov, V.V. (1981) On integral inequalities for trigonometric polynomials and their derivatives. Izv. **45**, 3–22; Ch. 4:104
Atkinson, K., Sharma, A. (1969) A partial characterization of poised Hermite-Birkhoff interpolation problems. SIAM Num. **6**, 230–235; Ch. 4:126, 130
Bajsanski, B., Bojanic, R. (1964) A note on approximation by Bernstein polynomials. BAMS **70**, 675–677; Ch. 10:311
Barrow, D.L., Chui, C.K., Smith, P.W., Ward, J.D. (1978) Unicity of best mean approximation. Math. Comp. **32**, 1131–1143; Ch. 12:389
Bartlet, M.W. (1975) Weak Chebyshev sets and splines. JAT **14**, 30–37; Ch. 3:94
Bavinck, H. (1976) Approximation processes for Fourier-Jacobi expansions. Appl. Anal. **5**, 293–312; Ch. 10:332

Becker, M. (1979) An elementary proof of inverse theorems for Bernstein polynomials. Aequationes Math. **19**, 145–150; Ch. 10:315
Becker, M., Nessel, R.J. (1978) A global approximation theorem for Meyer-König and Zeller operators. Math. Z. **160**, 195–206; Ch. 10:331
Berens, H. (1964) Approximationssätze für Halbgruppenoperatoren in intermediären Räumen. Schr. Math. Inst. Univ. Münster **32**; Ch. 9:295
Berens, H., DeVore, R.A. (1976) Quantitative Korovkin theorems for L_p-spaces. In: G.G. Lorentz, et al. (eds.) Approximation Theory II. Academic Press, New York, pp. 285–298; Ch. 9:301
Berens, H., DeVore, R.A. (1978) Quantitative Korovkin theorems for positive linear operators on L_p-spaces. TAMS **245**, 349–361; Ch. 9:301
Berens, H., DeVore, R.A. (1980) A characterization of Bernstein polynomials. In: E.W. Cheney (ed.) Approximation III. Academic Press, New York, pp. 213–219; Ch. 10:331
Berens, H., Lorentz, G.G. (1972) Inverse theorems for Bernstein polynomials. Indiana University Math. J. **21**, 693–708; Ch. 10:312, 313, 322
Berman, D.L. (1952) On a class of linear operators. Dokl. **85**, 13–16; Ch. 9:282
Besov, O.V. (1961) Investigation of a family of function spaces in connection with theorems of imbedding and extension. Trudy Mat. Inst. Steklov **60**, 42–81; Ch. 2:46
Birkhoff, G.D. (1906) General mean value and remainder theorems with applications to mechanical differentiation and quadrature. TAMS **7**, 107–136; Ch. 4:124
Birman, M.S., Solomyak, M. (1967) Piecewise polynomial approximation of functions of the class W_p^α. Mat. Sb. **2**, 295–317; Ch. 12:368, 389
Bishop, E. (1961) A generalization of the Stone-Weierstrass theorem. Pacific J. Math. **11**, 777–783; Ch. 1:14
Blichfeldt, H.F. (1901) Note on the functions of the form $f(x) \equiv \phi(x) + a_1x^{n-1} + \cdots + a_n$. TAMS **2**, 100–102; Ch. 3:58
Bojanov, B.D. (1982) An extension of the Markov inequality. JAT **35**, 181–190; Ch. 4:132
Bojanic, R., DeVore, R.A. (1969) A proof of Jackson's theorem. BAMS **75**, 364–367; Ch. 9:275
de Boor, C. (1963) Best approximation properties of spline functions of odd degree. J. Math. and Mech. **12**, 747–749; Ch. 5:152
de Boor, C. (1968) On the convergence of odd degree spline interpolation. JAT **1**, 452–463; Ch. 13:395
de Boor, C. (1972) On calculating with B-splines. JAT **6**, 50–62; Ch. 5:138
de Boor, C. (1973) Good approximation by splines with variable knots. In: A. Meir, A. Sharma (eds.) Spline Functions and Approximation. Birkäuser, Basel, pp. 57–72; Ch. 12:367
de Boor, C. (1975) On bounding spline interpolation. JAT **14**, 191–203; Ch. 13:393, 394, 396, 431
de Boor, C. (1976_2) A bound on the L_∞-norm of the L_2-approximation by splines in terms of a global mesh ratio. Math. Comp. **30**, 765–771; Ch. 13:403, 406, 431
de Boor, C. (1976_3) Odd degree spline interpolation at a bi-infinite knot sequence. In: K. Scherer, R. Schaback (eds.), Approximation Theory, Bonn, 1976. Lecture Notes in Mathematics, vol. 556. Springer, Berlin, pp. 30–53; Ch. 13:408
de Boor, C. (1987) The polynomials in the linear span of integer translates of a compactly supported function. CA **3**, 199–208; Ch. 13:417
de Boor, C., DeVore, R. (1985) A geometric proof of total positivity for spline interpolation. Math. Comp. **45**, 497–504; Ch. 5:165, 169
de Boor, C., Fix, G.J. (1973) Spline approximation by quasi-interpolants. JAT **8**, 19–45; Ch. 5:141, 142
de Boor, C., Jia, R.Q. (1985) Controlled approximation and characterization of the local approximation order. PAMS **95**, 547–553; Ch. 13:432

de Boor, C., Jia, R.Q., Pinkus, A. (1982) Structure of invertible (bi)infinite totally positive matrices. Linear Algebra and Applications **47**, 41–55; Ch. 13:431

de Boor, C., Pinkus, A. (1978) Proof of the conjectures of Bernstein and Erdös concerning the optimal nodes for polynomial interpolation. JAT **24**, 289–303; Ch. 9:301

de Boor, C., Rice, J. (1979) An adaptive algorithm for multivariate approximation giving optimal convergence rates. JAT **25**, 337–359; Ch. 12:389

Borosh, I., Chui, C.K., Smith, P.W. (1977) Best uniform approximation from a collection of subspaces. Math. Z. **156**, 13–18; Ch. 11:351

Brosowski, B., Deutsch, F. (1981) An elementary proof of the Stone-Weierstrass theorem. PAMS **81**, 89–92; Ch. 1:11

Brudnyi, Yu.A. (1959) Approximation by entire functions on the exterior of an interval and of the semi-axis. Dokl. **124**, 739–742; Ch. 8:240, 241

Brudnyi, Yu.A. (1963) Generalization of a theorem of A.F. Timan. Dokl. **148**, 1237–1240; Ch. 8:238

Brudnyi, Ju.A., Kruglyak, N.Ja. (1981) Real interpolation functors. Dokl. **23**, 5–8; Ch. 6:200

Burchard, H. (1974) Splines with optimal knots are better. J. Applicable Math. **3**, 309–319; Ch. 12:364, 366

Burchard, H. (1977) On the degree of convergence of piecewise polynomial approximation on optimal meshes II. TAMS **234**, 531–557; Ch. 12:383

Burchard, H.G., Hale, D.F. (1975) Piecewise polynomial approximation on optimal meshes. JAT **14**, 128–147; Ch. 7:228

Butler, G., Richards, F. (1972) An L_p saturation theorem for splines. Canadian Math. J. **24**, 957–966; Ch. 12:356

Butzer, P.L., Scherer, K. (1972) Jackson and Bernstein-type inequalities for families of commutative operators in Banach spaces. JAT **5**, 308–342; Ch. 9:295

Butzer, P.L., Stens, R.L., Wehrens, M. (1979) Approximation by algebraic convolution integrals. In: J.B. Prolla (ed.) Theory of Approximation and Functional Analysis. North Holland Publ. Co., 71–120; Ch. 8:239

Calderón, A.P. (1964) Intermediate spaces and interpolation, the complex method. Studia Math. **24**, 113–190; Ch. 6:200

Calderón, A.P. (1966) Spaces between L^1 and L^∞ and the theorem of Marcinkiewicz. Studia Math. **26**, 273–299; Ch. 2:57; Ch. 6:200

Carleson, L. (1964) Mergelyan's theorem on uniform polynomial approximation. Math. Scand. **15**, 167–175; Ch. 1:15

Cavaretta, A.S. (1974) An elementary proof of Kolmogorov's theorem. Amer. Math. Monthly **81**, 480–486; Ch. 5:154

Chaikan, G.M. (1974) An algorithm for high speed curve generation. Computer Graphics and Image Processing **3**, 346–349; Ch. 5:168

Cheney, E.W., Hobby, C.R., Morris, P.D., Schurer, F., Wulbert, D.E. (1969) On the minimal property of the Fourier projection. TAMS **143**, 249–258; Ch. 9:301

Ciesielski, Z. (1963) Properties of orthonormal Franklin systems. Studia Math. **23**, 141–157; Ch. 13:431

Ciesielski, Z. (1973) Constructive function theory and spline systems. Studia Math. **52**, 277–302; Ch. 12:359, 360

Clarkson, J.A., Erdös, P. (1943) Approximation by polynomials. Duke Math. J. **10**, 5–11; Ch. 11:348

Cox, M.G. (1980) The numerical evaluation of B-splines. National Physics Laboratory, Teddington, England; Ch. 5:138

Curry, H.B., Schoenberg, I.J. (1966) On Pólya frequency functions IV: The fundamental spline functions and their limits. J. d'Analyse Math. **17**, 71–107; Ch. 5:137, 141

Cwikel, M. (1976) Monotonicity properties of interpolation spaces. Ark. Mat. **14**, 213–236; Ch. 6:200

Czipszer, J., Freud, G. (1958) Sur l'approximation d'une fonction périodique et de ses dérivées successives par un polynome trigonométrique et par ses dérivées successives. Acta Math. **99**, 33–51; Ch. 7:207, 237

Dahmen, W., Görlich, E. (1974) Asymptotically optimal linear approximation processes and a conjecture of Golumb. In: P.L. Butzer, B. Sz.-Nagy (eds.) Linear Operators and Approximation II, ISNM vol. 25. Birkhäuser, Basel, 327–335; Ch. 7:237

Daubechies, I. (1988) Orthonormal bases of compactly supported wavelets. Comm. on Pure & Applied Math. **41**, 909–996; Ch. 13:432

Davie, A.M. (1975) The Banach approximation problem. JAT **13**, 392–394; Ch. 9:267

Demko, S. (1977) Inverses of band matrices and local convergence of spline projections. SIAM Num. **14**, 141–157; Ch. 13:404

Demko, S. (1985) On the existence of interpolation projectors onto spline spaces. JAT **43**, 151–156; Ch. 13:396

Derrienic, M.M. (1981) Sur l'approximation de functions intégrables sur $[0,1]$ par des polynômes de Bernstein modifies. JAT **31**, 325–343; Ch. 10:327

DeVore, R.A. (1968) L_p approximation of functions whose r-th derivative is of bounded variation. Studia Sci. Math. Hung. **3**, 443–450; Ch. 3:95

DeVore, R.A. (1976) Degree of approximation. In: G.G. Lorentz, et al. (eds.) Approximation theory II. Academic Press, New York, 117–161; Ch. 8:250, 252, 255, 266

DeVore, R.A. (1987) A note on adaptive approximation. Approx. Theory and Appl. **3**, 74–78; Ch. 12:389

DeVore, R.A., Jawerth, B., Popov, V. (1992) Compression of wavelet decompositions. Amer. J. Math. **114**, 737–785; Ch. 13:432

DeVore, R.A., Kierstead, H., Lorentz, G.G. (1987) A proof of Borsuk's theorem. In: Functional analysis, E. Odell and H. Rosenthal (eds.), Lecture Notes in Mathematics, vol. 1332. Springer, Berlin, 195–202; Ch. 3:85

DeVore, R.A., Popov, V. (1988_1) Interpolation spaces and non-linear approximation. In: Function spaces and applications, M. Cwikel et al. (eds.), Lecture Notes in Mathematics, vol. 1302. Springer, Berlin 1988, 191–205; Ch. 12:388

DeVore, R.A., Popov, V. (1988_2) Interpolation of Besov spaces. TAMS **305**, 297–314; Ch. 12:381

DeVore, R.A., Richards, F. (1972) Saturation and inverse theorems for spline approximation. In: Spline functions and approximation theory, A. Meir, A. Sharma (eds.) ISNM 21. Birkhäuser, Basel, 73–82; Ch. 12:356, 389

DeVore, R.A., Riemenschneider, S.D., Sharpley, R. (1979) Weak interpolation in Banach spaces. J. Functional Anal. **33**, 58–91; Ch. 6:181

DeVore, R.A., Scherer, K. (1979) Interpolation of linear operators on Sobolev spaces. Ann. Math. **109**, 583–589; Ch. 6:199, 200

DeVore, R.A., Scherer, K. (1980) Variable knot, variable degree spline approximation to x^β. In: Quantitative approximation, R. DeVore, K. Scherer (eds.). Academic Press, New York, 121–131; Ch. 12:389

DeVore, R.A., Scott, L.R. (1984) Error bounds for Gaussian quadrature and weighted L^1-polynomial approximation. SIAM Num. **21**, 400–412; Ch. 8:260

DeVore, R.A., Yan, Z. (1986) Error analysis for piecewise quadratic curve fitting algorithms. Computer Aided Geometric Design **3**, 205–215; Ch. 13:428, 429

Ditzian, Z. (1980_1) On interpolation of $L_p[a,b]$ and weighted Sobolev spaces. Pacific J. Math. **90**, 307–323; Ch. 10:322

Ditzian, Z. (1988) On the Marchaud-type inequality. PLMS **103**, 198–202; Ch. 2:49

Ditzian, Z., Ivanov, K.G. (1991) Strong converse inequalities. J. d'Analyse (to appear); Ch. 10:322, 327

Douglas, J., Dupont, T., Wahlbin, L. (1975) Optimal L_∞ error estimates for Galerkin approximations to solutions to two point boundary value problems. Math. Comp. **29**, 475–483; Ch. 13:406

Dyn, N., Lorentz, G.G., Riemenschneider, S.D. (1982) Continuity of the Birkhoff interpolation. SIAM Num. **19**, 507–509; Ch. 4:119

Dzyadyk, V.K. (1958) A further strengthening of the theorem of Jackson about the approximation by ordinary polynomials of continuous functions. Dokl. **121**, 403–406; Ch. 8:247, 252

Ehlich, H., Zeller, K. (1966) Auswertung der Normen von Interpolationsoperatoren. Math. Ann. **164**, 105–112; Ch. 9:271, 301

Eisenstadt, S.C., Jackson, K.R., Lewis, J.W. (1985) The order of monotone piecewise cubic interpolation. SIAM Num. **22**, 1220–1237; Ch. 13:424, 431, 433

Erdélyi, T. (1989) Markov type estimates for certain classes of constrained polynomials. CA **5**, 347–356; Ch. 4:132

Erdélyi, T. (1991) Estimates for the Lorentz degree of polynomials. JAT **67**, 187–198; Ch. 4:132

Erdélyi, T., Máté, A. and Nevai, P. (1992) Inequalities for generalized non-negative polynomials. CA **8**, 241–255; Ch. 4:132

Erdélyi, T., Szabados, J. (1988) On polynomials with positive coefficients. JAT **54**, 107–122; Ch. 4:132

Erdös, P. (1940) Extremal properties of polynomials. Ann. Math. **41**, 310–313; Ch. 4:112

Erdös, P., Turán, P. (1937) On interpolation I. Ann. Math. **38**, 142–155; Ch. 4:117

Erdös P., Vértesi, P. (1981) On the almost everywhere divergence of Lagrange interpolatory polynomials for arbitrary system of nodes. Acta Math. Acad. Sci. Hungar. **36** (1980) 71–89, **38** (1981) 263; Ch. 4:116

Esseen, C.G. (1960) Über die asymptotisch beste Approximation stetiger Funktionen mit Hilfe von Bernstein-Polynomen. Numer. Math. **2**, 206–213; Ch. 10:331

Favard, J. (1937) Sur les meilleures procédés d'approximation de certaines classes des fonctions far des polynômes trigonométriques. Bull. Sci. Math. **61**, I: 209–224; II: 243–256; Ch. 7:211, 213, 237

Fenchel, W. (1949) On conjugate convex functions. Canad. J. Math. **1**, 73–77; Ch. 3:62

Ferguson, D.R. (1969) The question of uniqueness for G.D. Birkhoff interpolation problems. JAT **2**, 1–28; Ch. 4:126, 130

Fiedler, H., Jurkat, W.B. (1983) Best L^1-approximation by polynomials. JAT **37**, 269–292; Ch. 9:302

Fisher, S.D., Morris, P.D., Wulbert, D.E. (1981) Unique minimality of Fourier projections. TAMS **265**, 235–246; Ch. 9:301

Freud, G. (1958) Eine Ungleichung für die Tschebyscheffsche Approximationspolynome. Acta Sci. Math. Hungar. **19**, 162–164; Ch. 3:77

Freud, G. (1959) Über die Approximation reeler stetiger Funktionen durch gewöhnliche Polynome. Math. Ann. **137**, 17–25; Ch. 8:238, 252

Freud, G. (1968) On approximation by positive linear methods II. Studia Sci. Math. Hungar. **3**, 365–370; Ch. 7:217; Ch. 9:280; Ch. 10:308

Freud, G., Popov, V. (1969) On approximation by spline functions. In: G. Alexits, S.B. Stechkin (eds.) Proceedings Conference on Constructive Theory of Functions. Akademia Kiado, Budapest, 163–172; Ch. 12:366

Fritsch, M.N., Carlson, R.E. (1980) Monotone piecewise cubic interpolation. SIAM Num. **17**, 238–246; Ch. 13:422, 423

Gaier, D. (1970) Saturation bei Spline-Approximation und Quadratur. Numer. Math. **16**, 129–140; Ch. 12:358

Garkavi, A.L. (1960) On simultaneous approximation of a periodic function and of its derivatives by trigonometric polynomials. Izv. **24**, 103–128; Ch. 7:237

Golitschek, M. v. (1970) Erweiterungen der Approximationssätze von Jackson im Sinne von C. Müntz. JAT **3**, 72–86; Ch. 11:345, 347

Golitschek, M. v. (1989) Short proofs of the inequalities of Szegö, Markov and Zygmund. In: Approximation and function spaces, vol. 22, Z. Ciesielski (ed.). Banach Center Publ., Warszaw, pp. 165–168; Ch. 4:97

Golitschek, M. v., Lorentz, G.G. (1989) Bernstein inequalities in L_p for $0 \le p \le +\infty$. Rocky Mountain Math. J. **19**, 145–156; Ch. 4:104

Gopengauz, A.I. (1967) To a theorem of A.F. Timan about the approximation of a function by polynomials on a finite interval. Mat. Zametki **1**, 163–172; Ch. 8:250

Haar, A. (1918) Die Minkowskische Geometrie und die Annäherung stetiger Funktionen. Math. Ann. **78**, 294–311; Ch. 3:72

Haverkamp, R. (1978) Zum Nullstellensatz von Krein für Haarsche Räume. JAT **23**, 104–107; Ch. 3:81

Heilmann, M. (1988) L_p-saturation of some modified Bernstein operators. JAT **54**, 260–273; Ch. 10:331

Hobby, C.R., Rice, J.R. (1965) A moment problem in L_1-approximation. PAMS **16**, 665–670; Ch. 3:48; Ch. 3:85

Hoeffding, W. (1971) The L_1 norm of the approximation error for Bernstein-type polynomials. JAT **4**, 347–356; Ch. 10:330

Höllig, K. (1981) L_∞ boundedness of L_2-projections on splines for a geometric mesh. JAT **33**, 318–333; Ch. 13:432

Höllig, K., Scherer, K. (1980) An identity for the characteristic polynomials of the L_2 projection on splines for a geometric mesh. In: Approximation Theory III, E.W. Cheney (ed.). Academic Press, New York; Ch. 13:431

Holmstedt, T. (1970) Interpolation of quasi-normed spaces. Math. Scand. **26**, 177–199; Ch. 6:194

Ivanov, K.G. (1980) Direct and converse theorems for the best algebraic approximation in $C[-1,1]$ and $L_p[-1,1]$. Coll. Math. Soc., Janós Bolyai, **35** Functions, Series, Operators, Budapest, 675–682; Ch. 8:239

Ivanov, K. (1982) On Bernstein polynomials. C.R. Acad. Bulgar. Sci. **35** (7) 893–896; Ch. 10:325

Jackson, D. (1924) A general class of problems in approximation. Amer. J. Math. **46**, 215–234; Ch. 3:86

Jetter, K., Lorentz, G.G., Riemenschneider, S.D. (1983) Rolle theorem method in spline approximation. Analysis **3**, 1–37; Ch. 5:158, 167

Johnen, H. (1972) Inequalities connected with moduli of smoothness. Mat. Vesnik **3**, 389–303; Ch. 6:176, 177

Johnen, H., Scherer, K. (1976) On the equivalence of the K-functional and moduli of continuity and some applications. In: Constructive Theory of Functions of Several Variables. Lecture Notes in Mathematics, vol. 571. Springer, Berlin New York, 119–140; Ch. 6:176, 178

Jones, R.C., Karlovitz, L.A. (1970) Equioscilliation under nonuniqueness in the approximation of continuous functions. JAT **3**, 138–145; Ch. 3:91, 92; Ch. 10:330

Kadec, M.I., Snobar, M.G. (1971) Some functionals over a compact Minkowski space. Mat. Zametki **10**, 694–696; Ch. 9:294

Kahane, J. (1961) Teoria Constructiva de Functiones. Course Notes University de Buenos Aires; Ch. 12:365

Karlin, S. (1973) Some variational problems on certain Sobolev spaces and perfect splines. BAMS **79**, 124–128; Ch. 5:167

Karlin, S., Ziegler, Z. (1966) Tchebysheffian spline functions. SIAM Num. **3**, 514–543; Ch. 5:162

Karlin, S., Ziegler, Z. (1970) Iteration of positive approximation operators. JAT **3**, 310–339; Ch. 10:330

Kelisky, R.P., Rivlin, T.J. (1967) Iterates of Bernstein polynomials. Pacific J. Math. **21**, 511–520; Ch. 10:330

Kemperman, J.H.B., Lorentz, G.G. (1979) Bounds for polynomials with applications. Indag. Math. **88**, 13–26; Ch. 4:100

Kilgore, T.A. (1978) A characterization of the Lagrange interpolating projection with minimal Tchebycheff norm. JAT **24**, 273–288; Ch. 9:301

Kirchberger, P. (1902) Über Tschebyscheff'sche Annäherungsmethoden. Dissertation, Göttingen; Ch. 3:58

Kolmogorov, A.N. (1935) Zur Grössenordnung des Restgliedes Fourierscher Reihen differenzierbarer Funktionen. Ann. Math. **36**, 521–526; Ch. 11:334

Kolmogorov, A.N. (1939) On inequalities between the upper bounds of the successive derivatives of functions on an infinite interval. Uchenye Zap. MGU, Mat. **30** (3) (1939) 3–13 (AMS Transl. Ser. 1,2 (1962) 233–243); Ch. 5:153

Kolmogorov, A.N. (1948) A remark on the polynomials of Chebyshev deviating the least from a given function. Uspehi Mat. Nauk **3**, 216–221; Ch. 3:63

König, H., Tomczak, N. (1990) Bounds for projection constants and the 1-summing norms. TAMS July, 1990; Ch. 9:302

Konjagin, S.V. (1981) On a problem of Littlewood. Izv. **45** (1981) (Mathematics of the USSR, Isvestia **18** (1982) 205–225); Ch. 9:285

Korevaar, J. (1947) A characterization of the sub-manifold of $C[a,b]$ spanned by the sequence $\{x^{\lambda_k}\}$. Indag. Math. **9**, 360–368; Ch. 11:348

Korneicuk, N.P. (1961) The best uniform approximation of certain classes of continuous functions. Dokl. **141**, 304–307 (AMS Transl. **2**, 1254–1259); Ch. 6:175

Korovkin, P.P. (1957) On convergence of linear positive operators in the space of continuous functions. Dokl. **114**, 961–964; Ch. 1:8

Krein, M.G. (1938) The L-problem in an abstract linear normed space. Transl. Math. Monographs **2**, Amer. Math. Soc., Providence, 1962; Ch. 3:85

Krein, M.G. (1951) The ideas of P.L. Cebysev and A.A. Markov in the theory of limiting values of integrals and their further development. Uspehi Mat. Nauk **6**, 3–122 (AMS Trans. Ser.2 **12**, (1959) 1–122); Ch. 3:80

Landau, E. (1913) Einige Ungleichungen für zweimal differentierbare Functionen. Proc. London Math. Soc. (2) **13**, 43–49; Ch. 5:153

Lax, P.D. (1944) Proof of a conjecture of P. Erdös on the derivative of a polynomial. BAMS **50**, 509–513; Ch4:132

Lebed, G.K. (1957) Inequalities for polynomials and for their derivatives. Dokl. **117**, 570–572; Ch. 8:241

Leviatan, D. (1982) The behavior of the derivatives of the algebraic polynomials of best approximation. JAT **35**, 169–176; Ch. 8:248

Lewis, D.R. (1978) Finite dimensional subspaces of L_p. Studia Math. **58**, 207–212; Ch. 9:288, 293

Ligun, A.A. (1985) On exact constants in the inequalities of Jackson's type. Mat. Zametki **38**, 248–256; Ch. 11:353

Lőfström, J. (1985) Best approximation in $L_p(w)$ by algebraic polynomials. Studia Sci. Math. Hungar. **20**, 375–394; Ch. 8:239

Lorentz, G.G. (1937) Zur Theorie der Polynome von S. Bernstein. Mat. Sb. **2**, 543–556; Ch. 10:306

Lorentz, G.G. (1950) Some new functional spaces. Ann. Math. **51**, 37–55; Ch. 2:23

Lorentz, G.G. (1963) The degree of approximation by polynomials with positive coefficients. Math. Ann. **151**, 239–251; Ch. 4:109, 112

Lorentz, G.G. (1964) Inequalities and the saturation classes of Bernstein polynomials, In: On approximation theory, P.L. Butzer et al. (eds.). Birkhäuser, Basel, 200–207; Ch. 10:311

Lorentz, G.G. (1964) Problem, In: On approximation theory, P.L. Butzer et al. (eds.). Birkäuser, Basel, p. 185; Ch. 8:250

Lorentz, G.G. (1966) Metric entropy and approximation. BAMS **72**, 903–937; Ch. 7:234

Lorentz, G.G. (1975) Zeros of splines and Birkhoff's kernel. Math. Z. **142**, 173–180; Ch. 4:133; Ch. 5:158

Lorentz, G.G. (1984) Convergence theorems for polynomials with many zeros. Math. Z. **186**, 117–123; Ch. 4:101

Lorentz, G.G., Schumaker, L.L. (1972) Saturation of positive operators. JAT **5**, 413–424; Ch. 10:331

Lorentz, G.G., Shimogaki, T. (1971) Interpolation theorems for the pairs of spaces (L^p, L^∞) and (L^1, L^q). TAMS **159**, 207–221; Ch. 6:200

Lorentz, G.G., Tomczak-Jaegermann, N. (1984) Projections of minimal norm. Longhorn Notes, The University of Texas, Texas Functional Analysis Seminar, 1983-1984, 167–176; Ch. 9:288

Machado, S. (1959) On Bishop's generalization of the Stone-Weierstrass theorem. PAMS **10**, 822–824; Ch. 1:11

Maier, V. (1978) The L_1-saturation class of the Kantrovic operator. JAT **22**, 223–232; Ch. 10:315

Marchaud, A. (1927) Sur les dérivées et sur les differences des fonctions de variables réelles. J. Math. pures et appl. **6**, 337–425; Ch. 2:47

Marsden, M.J. (1974) Quadratic spline interpolation. BAMS **80**, 903–906; Ch. 13:395, 431

McAllister, D.F., Roulier, J.A. (1981) An algorithm for computing a shape preserving osculatory quadratic spline. ACM Trans. Math. Software **7**, 331–347; Ch. 13:428

McGehee, O.C., Pigno, L., Smith, B. (1981) Hardy's inequality and the L^1-norm of exponential sums. Ann. Math. **113**, 613–618; Ch. 9:285

Meir, A., Sharma, A. (1969) On uniform approximation by cubic splines. JAT **2**, 270–274; Ch. 13:400

Micchelli, C.A. (1976) Cardinal L-splines. In: Studies in spline functions and approximation theory, S. Karlin, C.A. Micchelli, A. Pinkus, I.J. Schoenberg (eds.). Academic Press, New York, 203–250; Ch. 13:431

Mitjagin, B.S., Semenov, E.M. (1977) Lack of interpolation of linear operators in spaces of smooth functions. Izv. **41**, 1289–1328 (AMS Transl. **11** (1977) 1229–1266); Ch. 6:175, 199

Mørken, K. (1984) On two topics in spline theory. Ph.D. Thesis, University of Oslo; Ch. 13:396

Müntz, C.H. (1914) Über den Approximationssatz von Weierstrass. H.A. Schwarz Festschrift, Berlin, 303–312; Ch. 11:345

Murnaghan, F.D., Wrench, J.W., Jr. (1959) The determination of the Chebyshev approximating polynomial for a differentiable function. Math. Tables and other Aids to Comp. **13**, 185–193; Ch. 3:96

Nemeth, A.B. (1966) Transformation of the Chebyshev systems. Math. Cluj **37**, 339–347; Ch. 4:128

Nevai, P. (1984) Mean convergence of Lagrange interpolation, III. TAMS **282**, 669–698; Ch. 4:117

Newman, D.J., Rivlin, T.J. (1976) Approximation of monomials by lower degree polynomials. Aequationes Math. **14**, 451–455; Ch. 11:352, 353

Newman, D., Shapiro, H. (1963) Some theorems on Chebysev approximation. Duke Math. J. **30**, 673–682; Ch. 3:77

Nikolskii, S.M. (1946) Approximation of functions by trigonometric polynomials in the mean. Izv. **10**, 207–256; Ch. 7:215, 219

Nikolskii, S.M. (1951) Inequalities for entire functions of finite order. Trudy Mat. Inst. Steklov **38**, 513–532; Ch. 4:102

O'Neil, R., Weiss, G. (1963) The Hilbert transform and rearrangement of functions. Studia Math. **23**, 189–198; Ch. 2:33

Oswald, P. (1978) Ungleichungen vom Jackson-Typ für die algebraische beste Approximation in L_p. JAT **23**, 113–136; Ch. 8:239, 266

Oswald, P. (1980) Spline approximation in L_p, $(0 < p < 1)$ metric. Math. Nachrichten **94**, 69–96; Ch. 12:384

Peetre, J. (1970) A new approach in interpolation spaces. Studia Math. **34**, 23–42; Ch. 6:196

Peetre, J., Sparr, G. (1972) Interpolation of normed Abelian groups. Ann. Mat. Pura Appl. **92**, 217–262; Ch. 6:200; Ch. 12:383

Petrushev, P. (1988) Direct and converse theorems for spline and rational approximation and Besov spaces. In: Functions spaces and approximation, M. Cwikel et al. (eds.), Lecture Notes in Mathematics, vol. 1302. Springer, Berlin, 363–377; Ch. 12:385, 387

Phillips, G.M. (1979) Error estimates for best polynomial approximation, In: Approximation theory, A. Talbot (ed.). Academic Press, London, 1–6; Ch. 7:222

Pinkus, A., Ziegler, Z. (1979) Interlacing properties of the zeros of the error functions in best L_p approximations. JAT **27**, 1–18; Ch. 3:96

Pólya, G. (1922) On the mean-value theorem corresponding to a given linear homogeneous differential equation. TAMS **24**, 312–324; Ch. 3:96

Popoviciu, T. (1935) Sur l'approximation des fonctions convexes d'ordre superieur. Math. Cluj **10**, 49–54; Ch. 10:308

Prautzsch, H. (1984) A short proof of the Oslo algorithm. Computer Aided Geometric Design **1**, 95–96; Ch. 5:167

Ransford, T.J. (1984) A short elementary proof of the Bishop-Stone Weierstrass theorem. Math. Proc. Camb. Phil. Soc. **96**, 309–311; Ch. 1:11

Remez, E.Ya (1934) Sur le calcul effectif des polynômes d'approximation de Tschebyscheff. C.P. Paris **199**, 337–340; Ch. 3:78

Riemenschneider, S.D. (1978) The L_p-saturation of the Kantorovic-Bernstein polynomials. JAT **23**, 158–162; Ch. 10:315

Rivlin, T.J., Shapiro, H.S. (1961) A unified approach to certain problems of approximation and minimization. SIAM Num. **9**, 670–699; Ch. 3:66

Scheick, J.T. (1966) Some problems in approximation theory. Ph.D. dissertation, Syracuse University; Ch. 4:109

Scheick, J.T. (1972) Inequalities for derivatives of polynomials of special type. JAT **6**, 354–358; Ch. 4:109

Scherer, K. (1974) Characterization of generalized Lipschitz classes by best approximation with splines. SIAM Num. **11**, 283–304; Ch. 12:356

Schoenberg, I.J. (1946) Contributions to the problem of approximation of equidistant data by analytic functions. Quarterly Applied Math. **4**, 45–99 (Part A), 112–141 (Part B); Ch. 13:415, 416

Schoenberg, I.J. (1959) On variation diminishing approximation methods. In: R.E. Langer (ed.) On Numerical Approximation. Univ. of Wisconsin Press, Madison, pp. 249–271; Ch. 10:308, 310

Schoenberg, I.J. (1964) On interpolation by spline functions and its minimal properties. In: On approximation theory, P.L. Butzer, J. Korevaar (eds.), ISNM vol. 5. Birkäuser, 109–129; Ch. 5:152

Schoenberg, I.J. (1966) On the Hermite-Birkhoff interpolation. J. Math. Anal. Appl. **16**, 538–543; Ch. 4:124, 133

Schoenberg, I.J., Whitney, A. (1953) On Pólya frequency functions III. The positivity of translation determinants with applications to the interpolation problem by spline curves. TAMS **74**, 246–259; Ch. 5:162

Schumaker, L.L. (1983) On shape preserving quadratic spline interpolation. SIAM Num. **20**, 854–864; Ch. 13:426

Sedaev, A.A., Semenov, E.M. (1971) On the possibility of describing interpolation spaces in terms of the Peetre K-functional. Optimizaciya **4**, 98–114; Ch. 6:200

Sendov, B. (1987) On the theorem and constants of H. Whitney. CA **3**, 1–12; Ch. 6:184

Shalashova, L.Ya (1972) Approximating in the mean functions defined on a finite interval. Sib. Math. Zh. **13**, 679–687; Ch. 12:374

Shnirelman, L.G. (1938) On uniform approximations. Izv. **2**, 53–60; Ch. 3:95

Shvedov, A.S. (1979) Jackson's theorem in L^p, $0 < p < 1$, for algebraic polynomials and the order of co-monotonic approximation. Mat. Zametki **25**, 107–117; Ch. 12:374

Sikkema, P.C. (1971) Der Wert einiger Konstanten in der Theorie der Approximation mit Bernstein-Polynomen. Numer. Math. **3**, 107–116; Ch. 10:331

Smith, P.W. (1978) An improvement theorem for Descartes systems. PAMS **70**, 26–30; Ch. 11:351

Stechkin, S.B. (1951) On the order of approximation of continuous functions. Izv. **15**, 219–242; Ch. 7:205

Stechkin, S.B. (1948) Generalization of some inequalities of S.N. Bernstein. Dokl. **60**, 1511–1514; Ch. 4:131

Stein, E.M. (1957) Functions of exponential type. Ann. Math. **65**, 582–592; Ch. 5:156

Stens, R.L. (1978) Habilitationschrift. Aachen; Ch. 8:239

Stone, M.H. (1948) The generalized Weierstrass approximation theorem. Math. Magazine **21**, 167–184; 237–254; Ch. 1:13

Storozhenko, E.A. (1977) On approximation by algebraic polynomials of functions in the class L_p, $0 < p < 1$. Izv. **41**, 652–662; Ch. 12:374

Strang, G., Fix, G. (1973) A Fourier analysis of the finite element variational method. In: Constructive aspects of functional analysis, G. Geymonat (ed.). C.I.M.E., II Ciclo 1971, 793–840; Ch. 13:415, 432

Subbotin, Yu. (1967) On the relation between finite differences and the corresponding derivatives. Trudy Steklov Inst. Math. **78** (1965), 25–42, (AMS Transl. (1967) 23–42); Ch. 13:414

Subbotin, Yu., Chernykh, N.I. (1970) The order of best spline approximation to some functions classes. Mat. Zametki **7**, 31–42; Ch. 12:366

Sunouchi, G. (1968) Derivatives of a polynomial of best approximation. Jber. DMV **70**, 165–166; Ch. 9:295

Sunouchi, G., Watari, C. (1958-59) On determination of the class of saturation in the theory of approximation of functions I, II. Proc. Japan Acad. **34** (1958) 477–481; Tohoku Math. J. (2) **17** (1959) 480–488; Ch. 11:338

Szegö, G. (1928) Über einen Satz des Herrn Serge Bernstein. Schriften der Königsberger Gelehrten Gessellschaft **5**, 59–70; Ch. 4:97

Telyakovskii, S.A. (1966) Two theorems on the approximation of functions by algebraic polynomials. Mat. Sb. **70**, 252–265; Ch. 8:250

Timan, A.F. (1951) Strengthening of the theorem of Jackson about best approximation of continuous functions by polynomials on a finite interval of the real axis. Dokl. **78**, 17–20; Ch. 8:245

Timan, M.F. (1958) Converse theorems of the constructive theory of functions in the spaces L_p. Mat. Sb. **46** (88), 125–132; Ch. 7:210

Totik, V. (1983_1) Approximation by Meyer-König and Zeller type operators. Math. Z. **182**, 425–446; Ch. 10:331

Totik, V. (1983_2) Uniform approximation by Szasz-Mirakjan type operators. Acta Math. Hungar. **41**, 291–307; Ch. 10:331

Totik, V. (1984) An interpolation theorem and applications to positive operators. Pacific J. Math. **111**, 447–481; Ch. 10:322

Totik, V. (1991) Approximation by Bernstein polynomials. Preprint, 1991; Ch. 10:322

Trigub, R.M. (1962) Approximation of functions by polynomials with integral coefficients. Izv. **26**, 261–280; Ch. 8:245

Turán, P. (1939) Über die Ableitung von Polynomen. Compositio Math. **7**, 89–95; Ch. 4:132

Tureckii, A.H. (1960) On classes of saturation for certain methods of summation of Fourier series of continuous functions. Uspekhi Math. Nauk **15**, 149–156; Ch. 11:338

Varga, R.S. (1968) On an extension of a result of S.N. Bernstein. JAT **1**, 176–179; Ch. 7:233
Varga, R.S., Carpenter, A.J. (1985) On the Bernstein conjecture in approximation theory. CA **1**, 333–348; Ch. 9:275
Voronovskaya, E. (1932) Determination de la forme asymptotique d'approximation des functions par les polynomes de M. Bernstein. Dokl. A, 79–85; Ch. 10:307
Walsh, J.L., Motzkin, T.S. (1959) Polynomials of best approximation on an interval. Proc. Nat. Acad. of Sci. **45**, 1523–1528; Ch. 3:84
Whitney, H. (1934) Analytic extensions of differentiable functions defined on closed sets. TAMS **36**, 63–85; Ch. 6:182
Whitney, H. (1957) On functions with n-th bounded differences. J. Math. pure et appl. **36**, 60–92; Ch. 6:183, 200
Wulbert, D. (1971) Uniqueness and differential characterization of approximations from manifolds of functions. Amer. J. Math. **18**, 350–356; Ch. 3:95
Yan, Z. (1987) Piecewise cubic curve fitting algorithm. Math. Comp. **49**, 203–213; Ch. 13:433
Zedek, M. (1965) Cayley's decomposition and Pólya's W-property of ordinary differential equations. Israel J. Math. **3**, 81–86; Ch. 3:96
Zhuk, V.V. (1967) Some exact constants for the uniform approximations of periodic functions. Dokl. **201**, 263–266; Ch. 11:353
Zygmund, A. (1945) Smooth functions. Duke Math. J. **12**, 47–76; Ch. 2:51

Index